Psychological Review

RICHARD L. SOLOMON, Editor
University of Pennsylvania

CONTENTS

(A large group of psychologists has kindly agreed to assist in the editing of manuscripts submitted to the *Psychological Review*. In each issue the Editor will list in the *Review* those editorial consultants who have helped him process manuscripts for that issue, so that their efforts may be gratefully acknowledged.)

THE PSYCHOLOGICAL REVIEW

LEARNING THEORY AND THE NEW "MENTAL CHEMISTRY"[1]

W. K. ESTES

Indiana University

Hovering in the background of our scientific enterprises is a question which we, as investigators, can afford neither to raise very often nor to overlook entirely. Namely, do we have good reason to believe that the general methods and working assumptions underlying our research can be counted on to lead in the long run to satisfactory interpretations of our phenomena? In the psychology of learning, there has been a division of labor, with the experimentalists doing the overlooking, and each theorist having his turn at bringing up this question along with his answer. The answer is, of course, in each instance that the particular theorist's tactics point down the high road while those of the opposition lead into dark and forbidding *culs-de-sac*.

For my own part, although my hands are not entirely clean in the matter of theorizing, I had been content until very recently to go along with the single-minded "learning experimentalist," assuming that the Lord will look after those who remember their control groups and mind their Ps and Fs. The incident which jarred me out of this comfortable way of life began as a simple (in fact, as will be seen, unusually simple) bit of experimentation with no philosophical overtones. The original purpose of the experimentation was to provide more cogent empirical support than had hitherto been available for one of the central concepts of learning theory.

Several recent reviewers (e.g., Deese, 1958; Estes, 1956; Kendler, 1959; Restle, 1959) have noted that despite the popular stereotype of "learning theory" as virtually synonymous with "controversy," there has steadily, although unobtrusively, accumulated a body of concepts and assumptions which command relatively wide agreement and which contribute motivation and direction to a great part of the research being done in the field of learning. At the center of this core of communality one finds a concept which represents the distillation of centuries of theorizing about learning, not to speak of 70-odd

years of experimentation in the tradition of functionalism and, later, behaviorism: the concept of associative strength. In the "mental mechanics" and "mental chemistry" of pre-experimental associationism, this concept was verbalized in terms of the strength of associations between ideas; with Thorndike, it became strength of stimulus-response bonds; with Hull, the basic quantitative constructs of habit strength and excitatory potential. And it is in terms of this concept that contemporary learning theorists express the basic distinction between learning and performance, as well as their fundamental postulates relating the growth of habits to the number of training trials ("reinforcements").

There is little disagreement even in the precise quantitative form of these postulates, perhaps because our standards of acceptability come directly from the observed forms of empirical curves relating probabilities of learned responses, as well as measures of resistance to extinction or forgetting, to the number of training trials. Thus in a whole array of contemporary theories we find exactly the same equations expressing the effect of a reinforced training trial upon associative strength (cf. Restle, 1959). In Hull's system, this equation is

$$\Delta H = k(M - H)$$

where ΔH represents the change in habit strength (H) on any reinforced trial, M is the maximum value of habit strength, and k is a constant. Spence's (1955) revision of Hull's theory is based on the same postulate, although there is some question as to whether the H in the equation should be replaced by E, representing excitatory potential (Spence, 1955, 1958). Similarly, in statistical learning theory (Estes, 1959) the corresponding assumption has been expressed in the form

$$\Delta p = \theta(1 - p)$$

where Δp represents the change in the proportion of stimulus elements connected to a given response, and θ is a constant (the proportion of stimulus elements sampled on the trial). In the stochastic model of Bush and Mosteller (1955), the same linear function is assumed, "strength" being simply identified with response probability. For application to learning theorists, a well-known quotation may aptly be reversed to read, "In strength there is union." And with respect to the experimental literature, it seems fair to say that the assumptions that associative strength increases with reinforcement, decreases during retention intervals, and generalizes to new stimuli are the cornerstones of most contemporary treatments of conditioning and elementary verbal learning.

Areas of substantial agreement concerning either facts or interpretations, let alone both, are rare enough in the field of learning so that one might think we would do well to treasure the one we have located and carefully avoid doing anything to disturb it. This I (along with most of my fellow theorists and experimentalists) was happy to do, until my hand was forced by some purely experimental developments. The first of these was the, by now well-known, work of Kimble and his associates (1955, 1956; Dufort & Kimble, 1958) who, working with an eyelid conditioning situation, tried the novel procedure of omitting the CS on a substantial block of trials during an acquisition series. The rather surprising result of this variation was that the course of acquisition was virtually unaffected. Their interpretation was that only the first few reinforcements are actually effective in modifying the strength of associations between CS and CR, the

rest of the "conditioning curve" actually reflecting some nonassociative process. The second development was contributed by Rock (1957), who found no retardation in the speed of paired-associate learning when he introduced the device of replacing missed items with new ones at the end of every trial. Rock interpreted his findings as indicating that associations form on an all-or-none basis and that it is only associative strength in the sense of resistance to forgetting that grows as a function of number of reinforcements.

It seemed curious that both of these deviations from the usual experimental paradigms had yielded results which are in some respects sharply at variance with the generally accepted conception of the acquisition process. Of course these findings might turn out to have special explanations that would leave the established conception undisturbed. But then again, they might not. In the case of the Kimble, Mann, and Dufort (1955) study, counterexplanations and counterexperiments have already begun to pour into the literature (see, e.g., Goodrich, Ross, & Wagner, 1957) with the result so far of complicating matters still further rather than of clearing anything up. One might argue that concepts and assumptions which have been supported by a large accumulation of experimental findings cannot be seriously threatened by one or two apparently negative results. But this argument is weakened by the observation that all of the empirical support for the conceptualization of learning in terms of associative strength is quite indirect in character.

The basic concepts and assumptions of learning theory are universally supposed to refer to states and processes of the individual organism. Yet the existing evidence for the assumption that associative strength is an increasing function of number of reinforcements comes from performance curves representing average response measures over groups of learners, or from measures of resistance to extinction or retention scores averaged over groups of learners having different values of the independent variable. Even the few bits of negative evidence are indirect, depending on performance curves obtained under deviations from the usual experimental paradigms but still representing changes in average scores over series of trials for groups of subjects (Ss). It would seem that if our basic conceptions are sound, it should be possible to cut through the web of group performance curves and obtain more direct and compelling evidence for the existence of the assumed states and processes in individual organisms. This, in any event, is what I set out to accomplish for the concept of associative strength in a series of experiments now to be reported.

ON THE DEFINITIONS OF "REINFORCEMENT," "TEST TRIAL," AND "LEARNING"

In standard human learning situations, "learning" is almost universally defined and measured in terms of a change in the probability, or frequency, with which a given stimulating situation evokes a response (or instances of a response class) that has been designated as "correct" by the experimenter. With one reservation, to be noted below, I shall follow this usage. But the situation is quite different with "reinforcement," the same term being used in at least two quite different senses by different investigators and thus promoting no end of confusion. My own habitual usage is the "neutral definition" (Hilgard, 1956, p. 409) which identifies reinforcement empirically with the operation that is supplied by the experimenter in order to produce learn-

ing, as defined above, in any given situation. In a paired-associate situation, the reinforcing operation is the paired presentation of the stimulus and response members of an item; in classical conditioning it is the paired presentation of CS and US; in verbal conditioning, the reinforcing operation for a given predictive response (e.g., predicting that the left light will appear) is the occurrence of the corresponding event (appearance of the left light)—in each case without regard to whether the S correctly anticipated the response member of the paired-associate item, gave a CR prior to occurrence of the US, or correctly predicted the event on the trial in question. The only property that different types of reinforcing operations are assumed to share is their common quantitative effect on the conditional probabilities of the possible alternative responses to the stimulating situation in which reinforcement occurs.

A narrower definition, favored especially by writers associated with a drive-reduction interpretation of reinforcement, would limit the term reinforcement to an operation that follows and is contingent upon the occurrence of the reinforced response on any trial. In this usage, reinforcement in paired-associate learning occurs only when the S has made a correct response in anticipation of the paired stimulus-response presentation, and reinforcement in verbal conditioning occurs only on trials when the S correctly predicts the trial outcome. Whether, according to this view, reinforcement occurs on only those trials of a classical conditioning experiment on which a CR occurs prior to the US depends upon theoretical decisions as to whether the CR and UR are "the same response" and whether reinforcement occurs at the onset or the termination of the US.

The primary advantage I see in the "neutral definition" is that it can be applied in an objective and consistent manner independently of one's position on systematic or theoretical issues. Learning certainly may occur prior to the first correct anticipation in a paired-associate experiment, prior to the first correct prediction in verbal conditioning, prior to the first CR in classical conditioning. The present usage permits us to speak, for example, about changes in probability of a response as a function of reinforcements on trials preceding its first occurrence, on the one hand, and changes as a function of reinforcements on trials including and following its first occurrence, on the other, without changing our definition of reinforcement.

It should be emphasized that the neutral definition does not beg such questions as whether presentation of a US following a CR constitutes the same reinforcing operation as presentation of a US on a trial when the CR did not occur; these two procedures represent instances of the same reinforcing operation if and only if they produce the same change in the probability of evocation of the CR by the CS. However, it seems strategic to avoid issues of this sort when, as in the present investigation, we are concerned with the nature of the changes in response tendencies during learning rather than with the conditions giving rise to these changes. Consequently, in the experiments to be reported, we have attempted so far as possible to avoid the customary confounding of reinforcement with antecedent response. In paired-associate situations, for example, we have deviated from the usual anticipation procedure by separating the reinforcement (paired-presentation of stimulus and response members of an item) from the test for learning (presentation of the stimulus member alone) so that an item may receive more than one reinforcement before the first test trial or may

receive repeated test trials without intervening reinforcement.

For purposes of measuring retention, it would be ideal if one could give test trials on which no learning at all occurred. Indeed, so long as "learning" is conceived solely in terms of the definition given above (increase in probability of the "correct" response to a given stimulus), this goal is not too difficult to approximate. It seems intuitively clear, and can be demonstrated empirically (Estes, Hopkins, & Crothers, 1960), that no systematic increase in probability of correct responses to, say, paired-associate items will occur over a series of trials on which the stimulus members are presented alone and the S's responses receive no reward or informational feedback from the experimenter. We cannot, however, rule out the possibility that on these trials there might be learning in the sense of an increase in probability of whatever responses, correct or incorrect, actually occur. In fact, there is evidence that such learning does occur, but at a relatively low rate compared to the learning that occurs on reinforced trials (Estes et al., 1960). Consequently, in the analyses to follow, I shall assume that unreinforced trials can be treated, without serious error, simply as "neutral" test trials when primary interest is in measuring the effects of preceding reinforced trials.

UNITARY ASSOCIATION VS. HABIT HIERARCHY

If the stimulus-response relation established by reinforcement were a unitary, all-or-none connection of some sort, then the learning of a new response to a stimulus would automatically displace a previously associated response. Contrariwise, in every variant of the concept of associative strength, it is assumed that a number of different responses may simultaneously be associated with the same stimulus, the relative strengths of association depending primarily upon preceding frequencies of reinforcement of the different responses in the presence of the stimulus. This latter assumption has been embodied in Hull's "habit family hierarchy."

To spell out the question at issue in experimental terms, let us suppose that first Response A and subsequently Response B have been reinforced in the presence of a given stimulus (or stimulus complex—the nature of the stimulation does not matter so long as it is the same on each trial) for each member of a group of Ss and that a test trial now reveals an observed probability (relative frequency) of .5 for each response. Does this mean that for each S both Response A and Response B now have approximately equal habit strengths and therefore equal probabilities of evocation by the given stimulus? Such is the interpretation required by the conception of associative strength (habit strength, excitatory potential, proportion of conditioned elements, or whatever) as well as by any of the contemporary models formulated in terms of continuously variable response probabilities for individual Ss. Stated in these concrete terms, however, it does not seem that the assumption need depend for support only on extremely indirect evidence, as has hitherto been the case. To test these implications of the strength concept quite directly, we apparently need only reinforce two different responses to the same stimulus for an individual S, and then, by means of a series of unreinforced test trials, give the two responses an opportunity to exhibit their relative strengths—if such exist. The predictions under test would be straightforward. On the hypothesis of associative strength, or habit family hierarchy, we expect individual

Ss to shift back and forth between the two previously reinforced responses over a series of tests. On the hypothesis of a unitary, all-or-none association between stimulus and response, we expect that an individual S who makes Response A, say, on the first test trial with a given stimulus will not shift to B on subsequent tests.

A paired-associate learning situation offered a number of convenient features for an empirical realization of the hypothetical experiment just described. The stimuli and reinforcements are readily controllable; and, by embedding a stimulus to which two different responses are to be reinforced in a conventional list, one can make S's task appear little different from familiar learning situations. For each of 20 Ss, the procedures were replicated with six eight-item lists, all made up on the same principles.[2] The stimuli were all consonant syllables and the responses one-syllable words. Half of the items in each list had single correct responses; examples are:

STIMULUS	RESPONSE
HTX	wish
JFR	sped

These were simply "ballast" and will not be considered further. The other half of the items had two correct responses for each stimulus; examples are:

STIMULUS	RESPONSES
DGR	thaw, weep
BCG	pink, rule

The two correct responses to each stimulus were reinforced equally often for each S. By "reinforcement," in this context, we mean simply a paired presentation of stimulus and correct response to S; by "test," a presentation of the stimulus member of an item

alone. In half of the replications, there were exactly two reinforcements preceding the test trials, one on each response; in the remaining replications, there were four reinforcements, two on each response, prior to the test trials. Following the reinforcements, a series of unreinforced test trials was given, the order of the stimuli being randomized anew for each test.

The critical data for our purposes are the proportions of cases in which the response given on the first test trial to a given stimulus was repeated on the second test. According to an interpretation in terms of associative strength, or habit family hierarchy, repetitions (AA or BB) and shifts (AB or BA) should have occurred with roughly equal frequencies.[3] According to the notion of a unitary association, AA and BB should have each occurred about half of the time and the shifts AB and BA not at all. The results were as follows: Following two reinforcements, there were 85% repetitions and 15% shifts on the first two test trials, the percentages being based on all items that had correct responses on both tests. Following four reinforcements, these values were 89% for repetitions and only 11% for shifts. On later pairs of tests, the frequencies of AB and BA shifts were even smaller. Thus the results do not offer very impressive support for the assumption that a habit hierarchy exists in the individual S following the reinforcement of two different responses to the same stimulus. On the other hand, the data appear quite harmonious with an assumption of unitary association, since the proportions of AB and BA shifts appear small

---

[2] This experiment was conducted at Indiana University with the assistance of E. J. Crothers.

[3] The response given on the first test trial was the more recently reinforced response in approximately 60% of cases. Thus more instances of AA than BB would be expected; nevertheless the proportions of repetitions and shifts expected on the associative strength hypothesis are nearly equal (.52:.48).

enough to be attributable to minor uncontrolled factors such, for example, as fluctuations in context (background stimulation, stimulus traces from preceding items) from one trial to the next.[4]

Although the present experiment failed to produce the anticipated direct support for the concept of associative strength, a "strength theorist," particularly one working within the framework of Hull's system, may not be too disturbed. With the benefit of an ad hoc assumption concerning the range of "behavioral oscillation," the model of Hull and Spence can be made to yield the prediction that, of two responses reinforced to the same stimulus, one or the other will dominate on test trials, thereby preventing shifts from one to the other on the part of individual Ss. In order to obtain a more decisive test of the strength conception, we evidently require an even simpler experiment in which only one response is ever reinforced.

With the hope of forestalling an indefinite regress through a series of progressively more refined experiments,

---

[4] It should be noted that these results cast doubt on the concept of habit hierarchy only as applied to response probabilities in the presence of a stimulus situation that has been manipulated as a unit over the series of reinforced and test trials. M. S. Schoeffler (personal communication) and, in another unpublished study, Crothers and I have obtained positive evidence for the existence of a habit hierarchy associated with a compound stimulus, components of which have been separately correlated with reinforcements during the training trials. The results with compounding provide a control for the present experiment. If the very high proportions of repetitions reported above were attributable to learning that occurred on the first test trial, or to some nonassociative variable, then a similar excess of repetitions should have occurred over successive tests with stimulus compounds; however, no such excess was observed either in Schoeffler's study of compounding or in ours.

suppose we ask what is the minimum set of operations and observations actually needed in order to demonstrate learning. Normally there must be a pretest in order to determine the initial probability of the to-be-learned behavior in the test situation; in practice the experimenter often has a priori information about initial response probabilities which makes the pretest dispensable. There must be a presentation of some reinforcing operation, and afterward a test to assess the change in performance produced by the reinforcement. If the function of response occurrences and nonoccurrences is to be determined, there will have to be a second test trial. And there we have it. Controlled comparisons relative to effects of the principal events occurring during an acquisition series can, in principle, be accomplished in an experiment running to about a trial and a half. By usual standards, this constitutes what can only be called a "miniature experiment." However, miniature experiments appeared to be what the tactical situation called for, and therefore miniature experiments are what we set out to run.

CONCEPTIONS OF THE ACQUISITION PROCESS: ASSOCIATIVE STRENGTH VS. ALL-OR-NONE MODELS

In the first of these experiments,[5] we used a paired-associate situation with consonant syllables as stimuli and numbers as responses. Forty-eight Ss were run with an eight-item list, yielding 384 observations on the first test trial. The principal portion of the experiment consisted simply in presenting each S once with each stimulus-response pair and then testing with each stimulus alone (in a new random order). Before pro-

---

[5] This experiment was conducted at Indiana University with the assistance of B. L. Hopkins; for a full report of the method and results see Estes et al. (1960).
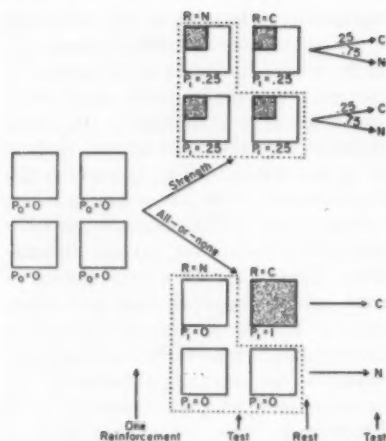
FIG. 1. Schema representing effects of a single reinforcement according to incremental (upper branch) vs. all-or-none (lower branch) theories. Squares represent Ss, with the proportion of darkened area in each indicating the probability of the correct response (C) for the given individual.

ceeding to the results, let us examine the outcome expected on the basis of the notion of learning as a change in associative strength. In Figure 1 the situation is schematized in terms of a single item. The four squares at the left represent four hypothetical Ss, the emptiness of the squares indicating that all start the experiment with zero probabilities of making the correct response. Now we give a single reinforcement (paired presentation of the stimulus and correct response), the result of which is to raise the probability of the correct response (C) to, say, .25. The upper arrow leads to the theoretical state of affairs after this reinforcement, according to an interpretation based on the conception of associative strength. The strength of the association is increased for all of the Ss; and, neglecting for the moment possible individual differences, the probability of the correct response is now .25 for each individual, the one at the upper right who

happened to make a correct response on the test, and the three who did not.

Suppose now that the interpretation based on the concept of strength were completely wrong and that stimulus-response associations really formed on an all-or-none basis. Then the state of affairs after the reinforcement should be as shown in the lower part of the figure. Again the probability of a correct response increases from zero to .25, but the .25 now refers only to the group, not to any individual S. One S has formed the association (darkened square), and three have been unaffected by the reinforcement (empty squares).

To distinguish empirically between these two logically possible outcomes, we need only add the remaining half-trial to our trial-and-a-half, i.e., give another test without intervening reinforcement. Now, if the upper branch of the diagram is essentially correct, all Ss should have equal probabilities of making the correct response on the second test trial, regardless of what they did on the first test. But if the lower branch is correct, correct responses on the second test should come only from Ss who made correct responses on the first test. None should come from Ss who made incorrect responses (N) on the first test, for these Ss would not have profited at all from the learning trial.

There might be some attenuation of the expected proportions of correct responses on Test 2 by Ss making correct responses on Test 1 if there is any forgetting in this situation, but the proportions of correct following incorrect provide a critical comparison. If the all-or-none view is correct, then this proportion should be zero, or at least no greater than could be achieved by sheer guessing. But if any version of the strength conception is correct, then the proportion of correct following incorrect responses should be greater than

chance. In order to make the outcomes that can be tolerated by the two interpretations sharply different, we need only choose our experimental materials and conditions so that the overall proportions correct on both first and second tests are well above chance. It can be seen in Figure 2 that this has been achieved, for approximately 50% of the items were correct on the first test and nearly 40% on the second. Considering the critical lower branch of the diagram, leading from an incorrect response on the first test to a correct response on the second, we see that the results lean strongly in the direction prescribed by an all-or-none conception, for the 9% of correct following incorrect responses is less than the 12½% that could be achieved even by rather unintelligent guessing with an eight-item list if the reinforcement had no effect at all on these items. The difference between this value and the 71% of correct following correct responses is so large that a statistical test would be an empty formality.

A possible defense that might be advanced by a "strength theorist" is the hypothesis that the 51% of cases with incorrect responses on the first test simply represent preponderantly slower learners or more difficult items than the 49% of cases with correct responses. If so, then a control condition in which a second reinforcement is given between the first and second tests should yield a percentage of correct responses on Test 2 following incorrect on Test 1 that is much smaller than the percentage correct on Test 1. This control was run (with the same 48 Ss but different items), and the result is shown in Figure 3. The effect of the first reinforcement on the full set of Ss and items was to raise the probability of a correct response from near zero to .40; the effect of the second reinforcement on cases having incorrect responses on

the first test was to raise the probability of a correct response from near zero to .46. Thus there seems to be no support forthcoming for the hypothesis of a large difference in learning rate between cases which did and cases which did not have correct responses on the first test.

Although it would be nice to claim credit for rare prescience in predicting the outcome of this little experiment, the fact is that the result came as a distinct jar to my preconceptions. In designing the study, our idea was not to undermine the strongly entrenched concept of associative strength, but to support it by showing that the results of Rock's experiments, apparently calling for an all-or-none interpretation, must be attributed to some artifact concealed in his ingenious but somewhat complex procedures. Thus when Hopkins and I examined the data from our initial group of 24 Ss and found the pattern shown in Figures 2 and 3, our first reaction was to replicate the whole thing with another group. But when the two replications turned out to agree in every essential respect, we were left
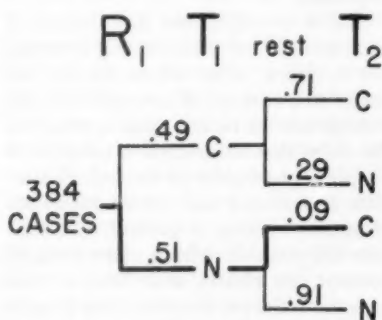

FIG. 2. Results of miniature experiment on acquisition of paired-associates. Empirical values are proportions of instance in which correct (C) and incorrect (N) responses on first test trial after a single reinforcement (paired presentation of stimulus and response members) were followed by C and N responses on a second test trial.

W. K. Estes

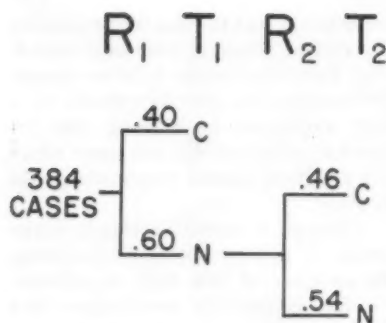$$R_1 \quad T_1 \quad R_2 \quad T_2$$

FIG. 3. Proportion correct on test after a second reinforcement for cases not having correct responses on first test compared with proportion correct on first test for the full set of Ss and items. The Ss and situation are the same as those represented in Figure 2.

with no obvious course but to begin digesting an unanticipated and not entirely palatable conclusion. The most cleanly controlled comparisons we had managed to devise yielded no evidence that repeated reinforcements in this situation have any function other than to give repeated opportunities for the discontinuous formation of a learned association between observed stimuli and responses.

Still, it is well known that theoretical doctrines do not yield readily to negative evidence. One whose theories are based on a concept of strength will lose little ground if he can make a stand on the claim that all-or-none acquisition is simply a peculiarity of the paired-associate experiment and not characteristic of human learning in general. To evaluate this possible defense of the strength concept, we clearly shall have to turn to some different situation that is quite different in the response mechanism and reinforcing operations from paired-associate learning. Eyelid conditioning meets these specifications, and it is convenient for our purposes since a colleague, I. Gormezano, has kindly made available his data from an in-

tensive period of data collecting in the Wisconsin conditioning laboratory. Gormezano trained a sufficiently large group, approximately 170 Ss, under identical conditions so that the first few acquisition trials can be treated as one of our miniature experiments and analyzed in much the same way as the paired-associate study.

The situation obtaining over the first couple of trials is schematized in Figure 4. In the diagram, T_1 is the first CS presentation, prior to the first reinforcement, and we shall consider only Ss who made no CR on this test. Thus the initial probability of a CR is taken to be zero. Suppose now that the effect of the first reinforcement is to raise the probability of a CR to .25. According to the strength conception, shown in the upper panel, each S has his strength of conditioning increased by the same amount by this reinforcement and now has probability .25 of making a CR. Then the second reinforcement increases the conditioned strength for each S again; and, regardless of whether or not a particular S happened to make a CR on T_2, he now
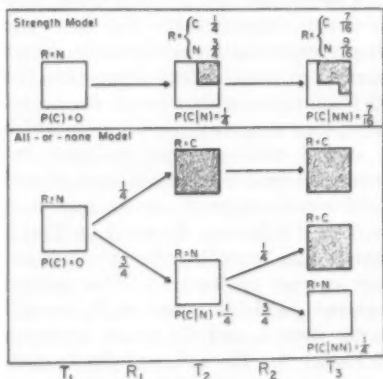

FIG. 4. Schema for first two trials of eyelid conditioning experiment showing changes in CR probability (proportion of darkened area in squares representing Ss) prescribed by incremental vs. all-or-none theories.

has a higher probability ($\frac{7}{16}$ if we apply the linear function mentioned earlier).[6] According to an all-or-none conception, the situation after the first reinforcement, shown in the lower panel, is that for $\frac{1}{4}$ of the Ss the CR has become associated with the CS and for the remaining $\frac{3}{4}$ of the Ss no conditioning has occurred. The effect of the second reinforcement is to give the unconditioned Ss another chance, and $\frac{1}{4}$ of these now become conditioned. The differential prediction, then, concerns the probability of a CR on the third test for Ss who made no CR on the second test (and similarly the probability of a CR on the fourth test for Ss who made none on any previous test, and so on). The strength conception requires this conditional probability to increase, whereas the all-or-none conception requires it to remain constant. The test seems quite sharp, for even with allowance for variation in conditioning rates among Ss, a model which assumes that associative strength increases with reinforcements cannot stand constancy of this probability unless its assumptions are so restricted that it reduces to an all-or-none model.

The pertinent results of Gormezano's study are shown in Figure 5, carried through the first four trials, beyond which the number of cases begins to drop off too much for comfort. Inspecting the sequence of probabilities of CRs after 1, 2, or 3 consecutive NCRs—.13, .15, .14—we find the hypothesis of constancy appearing rather more attractive than the progressively

[6] For this example, the parameter θ in the function $\Delta p = \theta (1 - p)$ is equal to $\frac{1}{4}$ and after the first experiment p is also equal to $\frac{1}{4}$. Therefore we have

$$\Delta p = \frac{1}{4}(1 - \frac{1}{4}) = \frac{3}{16}$$

and for the new probability after the second reinforcement,

$$p + \Delta p = \frac{1}{4} + \frac{3}{16} = \frac{7}{16}$$

$$T_1\ R_1\ T_2\ R_2\ T_3\ R_3\ T_4$$
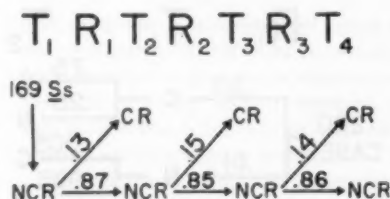
Fig. 5. Trial-by-trial acquisition data from Gormezano's study of eyelid conditioning. Values of particular interest are the proportions of CRs following 1, 2, or 3 consecutive non-CRs.

increasing trend required by the strength interpretation. (According to a linear model, for example, the value .15 for a CR after an NCR should have been .24, and the .14 for a CR after two NCRs should have been .34.)

The consistency of these conditioning data with those of the paired-associate situation is almost too good to be true. In psychology we are not used to having quantitative tests of alternative theoretical notions yield such apparently decisive outcomes. Consequently, and considering the importance of the theoretical issue, perhaps we will not yet be accused of beating a dead hypothesis if we look for one more test with experimental arrangements differing from both of those preceding. We would like a situation similar to paired-associate learning in that unreinforced test trials can readily be given without disturbing the learners but one which eliminates the possibility of achieving substantial proportions of correct responses by guessing. A situation which meets these desiderata is the free verbal recall experiment used by Bruner, Miller, and Zimmerman (1955). For our present purposes the minimal experiment will consist of a single reinforcement followed by two recall tests. The reinforcement involves merely the experimenter's reading a list of words aloud to S. On a recall test, S is asked to write down as many words as he can

R_1 T_1 T_2

280 CASES

.39 C — .75 C / .25 N

.61 N — .005 C / .995 N

816 CASES

.28 C — .54 C / .46 N

.72 N — .02 C / .98 N

Fig. 6. Results of two miniature experiments on free verbal recall, showing near-zero proportions of correct responses on a second test trial for cases which did not have correct responses on the first test after a single reinforcement.

remember (in any order). Then after an interval during which no additional reinforcement is given, S is (unexpectedly) tested again.

Results of two experiments [7] of this sort are shown in Figure 6. The upper tree represents an experiment with 35 Ss, each given a list of eight words at R_1. On the first test, T_1, 61% of the 280 opportunities for correct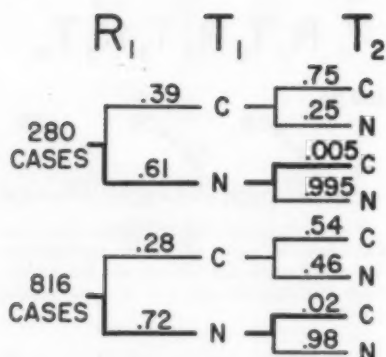 responses (C) yielded either incorrect responses or omissions (N), and of these less than 1% were followed by correct responses on the second test, T_2. In a replication conducted with some minor variations in procedure, 102 Ss were presented with eight words each on the reinforced trial. This time (lower tree in Figure 6) there were 72% N responses on the first test and less than 2% of these were followed by C responses on the second test. Clearly, if a word is not given correctly on the first test by a particular S, the chances are virtually *nil* that it will be correct on a second test.

[7] These experiments, conducted at Indiana University with the assistance of Judith Crooks, will be reported in detail elsewhere.

This result does not, of course, *prove* that reinforcement has exerted no strengthening effect on the associations in the cases when the correct response failed to occur on the first test. But one whose theory requires him to assume that such strengthenings occur has a taxing assignment in producing a case for the existence of factors or processes which appear in just sufficient force to cancel out the hypothesized increments in response strength under each set of experimental procedures we have examined. Explanations depending on such factors as individual differences are not very prepossessing in the light of control comparisons of the type exhibited in Figure 3 (a similar control, with a similar result, was used for the free-verbal-recall situation). One might appeal to the effects of learning which occurs on the first test trial itself, arguing that an incorrect response which occurs on the first test receives a large increment in associative strength (from sheer contiguity or perhaps from some unspecified source of reinforcement) and therefore recurs with high probability on the next test. One important difficulty with this hypothesis is that the data do not support it. In the paired-associate study cited above, for example, the observed relative frequency with which an incorrect response occurring on the first test was repeated on the second test was only .24. Interpretations which preserve the incremental conception of associative learning should certainly be sought with all vigor; at the time of writing, however, none has come to my attention that seems at all plausible.

RETENTION AS A FUNCTION OF REINFORCEMENT AND INTERSPERSED TEST TRIALS

The story does not end here. Even if one is ready to grant that associations are made and broken in an all-or-none

fashion, this concession, although by no means a minor one, does not exhaust the resources of the strength concept. It is possible that after an association has once formed, associative strength in the sense of resistance to forgetting continues to grow as a function of reinforcements. In fact the experimental situations we have used appear well suited to demonstrate such an effect.

It will not have escaped notice that although the three miniature experiments yielded sharply negative results relative to the conception of learning as a gradual strengthening of associations by reinforcement, they agree only in part with the usual notion of all-or-none learning (as, for example, defined by Voeks, 1955). The formation of learned associations appears to be discontinuous rather than gradual, but once a correct response (or conditioned response) has occurred to a given stimulus, it does not appear with certainty when the stimulus recurs after a rest interval. In fact, "forgetting" as measured by the proportion of shifts from a correct response on the first test to non-correct on the second test ranges from about 20% in the paired-associate situation to about 50% in free verbal recall. For the eyelid conditioning study, we did not report these proportions since the numbers of observations on which they were based were rather small; but for what it is worth, we might add that the proportions of CR to NCR shifts following the first three tests, despite the intervening reinforcement, fall well within the range given above for the other two experiments.

The minimum set of operations and observations required to test for a dependence of "resistance to forgetting" upon number of reinforcements is exhibited in Table 1. Following either one or two reinforcements, a sequence of two unreinforced test trials is given;

and retention is measured in terms of the proportion of instances in which a correct response to a given stimulus on the first of these tests is repeated on the second test. Results of two such experiments are shown in the table.

The left-hand column of proportions represents data from a study conducted by the writer with the assistance of E. J. Crothers (Estes et al., 1960). Twenty Ss were each run on six eight-item paired-associate lists under each condition (number of reinforcements). The stimuli were nonsense syllables, and the responses were numbers in half the lists and familiar words in the other half. The right-hand column presents data from the free verbal recall experiment previously cited (102 Ss each tested with a list of eight items under each condition). The pattern of test proportions needs little comment. In each case the difference called for by the traditional conception of associative strength fails to appear.

This outcome is a little hard to swallow. It is well known that retention increases with overlearning. If the additional reinforcements given during the overlearning period do not produce the increased resistance to forgetting, then what does? We may obtain a clue as to the answer by introducing what might a priori seem to be very slight change in the design of our minimal experiment on retention. This variable is illustrated in Table 2. In the free

TABLE 1

DESIGN AND RESULTS OF MINIMAL EXPERIMENT ON RETENTION AS A FUNCTION OF NUMBER OF REINFORCEMENTS

Procedure	Amount of Retention	
	P-A	FVR
R T$_1$ T$_2$	.90	.54
R R T$_1$ T$_2$	.89	.52

TABLE 2

DESIGN AND RESULTS OF MINIMAL EXPERI-
MENT ON RETENTION AS A FUNCTION
OF A TEST TRIAL INTERPOSED
BETWEEN REINFORCEMENTS

Procedure	Amount of Retention
R R T₁ T₂	.52
R T₁ R T₂ T₃	.78

verbal recall experiment cited previ-
ously, we included, in addition to the
condition shown in the upper row
(which is the same as in Table 1), a
condition with the same number of re-
inforcements but with an additional
interspersed test trial. The idea was
to give an opportunity for correct re-
sponses learned on the first reinforced
trial to become conditioned to cues
which are present only on test trials
and not on training trials—in other
words to spread the effect of the first
reinforcement over more stimuli. Re-
tention is measured in terms of the pro-
portion of correct responses repeated
from T_1 to T_2 in the upper row and
from T_2 to T_3 in the lower row. In
contrast to the lack of effect on reten-
tion produced by increasing the number
of reinforcements, note the large effect
of the added test trial. Two reinforce-
ments without an intervening test yield
only 52% retention, whereas two rein-
forcements with an intervening test trial
yield 78% retention from the first to
the second test following the second
reinforcement.

Although the principle that retention
increases with number of reinforce-
ments is exceedingly well established in
the lore of human learning, we must
face the possibility that this empirical
relation, like the classical acquisition
curve, is an artifact of the confoundings
inherent in the usual experimental par-
adigm.

CONCLUDING REMARKS

To recapitulate the box score: we
have conducted a series of highly sim-
plified experiments especially designed
to provide relatively direct evidence for
the widely accepted interpretation of
learning and retention in terms of incre-
ments and decrements in associative
strength. We noted that in virtually
all contemporary learning theories, the
concept of strength is assumed to have
at least three different empirical man-
ifestations—the habit hierarchy, the
growth of associative strength in the
sense of response probability as a func-
tion of number of reinforcements, and
the increase in resistance to forgetting
of once established associations with
additional reinforcement. We tested
for all three of these effects by means
of the simplest controlled comparisons
we could arrange, doing our utmost to
eliminate the confoundings and the lay-
ers of statistical processing that shield
the behavioral changes occurring on in-
dividual learning trials completely from
view in conventional experimental de-
signs. And under these presumably
favorable circumstances, all three ef-
fects mysteriously evaporated, leaving a
picture of unitary associations the learn-
ing and unlearning of which proceed on
an essentially all-or-none basis.[8]

[8] Several readers of a prepublication draft
of this article have raised the question
whether alternative measures, e.g., response
latencies or recognition scores, might yield
evidence of learning in cases where the prob-
abilities of conditioned responses or of cor-
rect recalls do not. Regardless of the an-
swer, it is important to note that information
about concomitant changes in other variables
would have no logical bearing upon conclu-
sions pertaining to the one actually chosen
for analysis in a given experiment. To
determine whether the behavior of recalling
the correct response to a paired-associate
item is learned in all-or-none fashion, we
required an analysis according to the par-
adigm of Figure 2 with recall score as the
dependent variable. To determine whether

What is the import of these results for the question raised earlier concerning the strategic soundness of the general methods on which investigators of learning chiefly rely? The concept we have examined in detail epitomizes the intervening-variable paradigm for theory construction that has been popularized by Tolman and Hull, and their followers, to the point of dominating contemporary learning theory. The general technique is to postulate a hypothetical state or entity which is held to intervene in some sense between observed stimulus and response variables. Predictions derived from theories built around such constructs are checked against data from standard learning experiments. Thus the adequacy of the theories depends in turn on the adequacy with which essential aspects of learning are captured by standard experimental designs.

I have emphasized the term "standard" for despite the fact, decried by many critics, that psychologists in this field have resisted all admonitions to standardize their tasks, situations, and procedures, they have more than compensated for this lack by the degree to which they have standardized their experimental designs. Scanning the reference list of, say, Hilgard's *Theories of Learning* (1956),[9] one can find no

---

correct recognition of a previously viewed item is learned on an all-or-none basis, we would need a similar analysis with recognition score as the dependent variable. To determine whether the behavioral change associated with a decrease in latency is learned on an all-or-none basis, we would need a similar analysis with some criterion of change in latency as the dependent variable, and so on.

[9] The harvest is even scantier if one consults the more monographic works, e.g., Tolman (1932), Hull (1943), or Spence (1955), associated with systems of intervening-variables in behavior theory; or, for that matter, my own theoretical writings (Estes, 1959). In the eclectic treatise of McGeoch and Irion

more than a meager handful of studies which are not molded into a paradigm that might be termed the groups-by-trials design. The master blueprint requires the investigator to average some type of performance score, e.g., frequency of conditioned responses or of correct responses, over a series of trials for groups of Ss—the groups being differentiated on the basis of conditions obtaining over the series. Going back a few years in the experimental journals, one finds the customary output of this experimental paradigm to be a set of mean performance curves for experimental and control groups; more recently the output is typically an array of analysis of variance tables, perhaps supplemented by some mean criterion scores. These observations in themselves do not constitute a criticism. No one would gainsay that the groups-by-trials design and its associated statistical techniques are useful tools for assaying the effects of various procedures and conditions upon performance. It is easy, however, to overlook the fact that the groups-by-trials design yields only information about relationships and trends which hold on the average over groups of Ss and series of trials. No accumulation of experiments, however large, all conducted and all analyzed in accord with this same general method can provide a sufficient empirical check on concepts and assumptions that refer to processes or events occurring in the individual learner. The findings we have considered in this paper suggest that in point of fact some of the most firmly entrenched concepts and principles of learning theory may be in a sense artifacts of a conventionalized methodology.

The laboratory investigator of learn-

(1952), deviant experimental designs appear in references with the earliest dates but disappear as one comes down to the contemporary literature.

ing is used to going his own way more or less oblivious to the rattle of criticisms from textbook writers, educators, and other "outsiders." Traditionally the criticisms have always been to the same effect—that the "learning experimentalist" should push on from the worn out fields of conditioning and simple verbal learning, where there are really no unsolved problems of any importance remaining, and devote himself to richer, more complex, experiments that come closer to learning situations of real life (as contrasted with the unreal life found in the laboratory). Here we break with tradition, for the criticism generated by our work within the field suggests that progress toward a satisfactory theory of learning requires, not more complex, but simpler experiments. The conventional experiment, far from being oversimplified, represents such a complex and intricate confounding of stimulus and response variables over trials that once it has been done and reported in conventional form, no amount of study of the analysis of variance tables and Vincentized performance curves can disclose the effects exerted by specific causal variables on individual Ss upon particular occasions. Concepts depending solely upon the conventional experiment for support may turn out to belong, not to a psychology of learning, but only to a psychology of the criterion score, the mean performance curve, and the groups-by-trials design.

In concluding, I would like to indicate that I do not mean to offer the experimental findings reported in this paper as a crucial test of incremental vs. all-or-none theory. What does seem clear with respect to the former is that the kind of evidence heretofore adduced in support of incremental theories is inadequate to distinguish them from alternative conceptions. More penetrat-

ing experimental analyses are required. We have made a start in this direction, and from the early returns it appears that no extant theory of the incremental type can handle the pattern of results that is emerging. The temptation is great to indulge now in a bit of speculation as to whether the answer to this situation will prove to be a remodeling of one of the familiar incremental theories, or perhaps a quite different theory based on all-or-none assumptions—but this is a step I do not intend to take.

Since the earliest days of associationism, overdependence upon speculation and circuitous inference has impeded the interplay of theory with experiment. While I would not for a moment depreciate the role of imagination in science, I suspect that it will begin to serve us effectively in learning theory only as we begin to accumulate reliable determinations of the effects of single variables upon single learning trials in individual organisms. If by continual simplification of our experimental analyses and refinement of our mensurational procedures we can achieve these determinations, we may find that the long sought laws of association may be not merely "instigated," or even "suggested," but literally dictated in form by empirical data.

REFERENCES

BRUNER, J. S., MILLER, G. A., & ZIMMERMAN, C. Discriminative skill and discriminative matching in perceptual recognition. *J. exp. Psychol.*, 1955, **49**, 187–192.

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning.* New York: Wiley, 1955.

DEESE, J. *The psychology of learning.* (Rev. ed.) New York: McGraw-Hill, 1958.

DUFORT, R. H., & KIMBLE, G. A. Ready signals and the effect of interpolated UCS presentations in eyelid conditioning. *J. exp. Psychol.*, 1958, **56**, 1–7.

ESTES, W. K. Learning. *Annu. Rev. Psychol.*, 1956, **7**, 1–38.

ESTES, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 380–491.

ESTES, W. K., HOPKINS, B. L., & CROTHERS, E. J. All-or-none and conservation effects in the learning and retention of paired associates. *J. exp. Psychol.,* 1960, in press.

GOODRICH, K. P., ROSS, L. E., & WAGNER, A. R. Performance in eyelid conditioning following interpolated presentations of the UCS. *J. exp. Psychol.,* 1957, **53,** 214–17.

HILGARD, E. R. *Theories of learning.* (Rev. ed.) New York: Appleton-Century-Crofts, 1956.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

KENDLER, H. H. Learning. *Annu. Rev. Psychol.,* 1959, **10,** 43–88.

KIMBLE, G. A., & DUFORT, R. H. The associative factor in eyelid conditioning. *J. exp. Psychol.,* 1956, **52,** 386–91.

KIMBLE, G. A., MANN, L. I., & DUFORT, R. H. Classical and instrumental eyelid conditioning. *J. exp. Psychol.,* 1955, **49,** 407–17.

McGEOCH, J. A., & IRION, A. L. *The psychology of human learning.* (Rev. ed.) New York: Longmans, 1952.

RESTLE, F. A survey and classification of learning models. In R. R. Bush and W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford, Calif.: Stanford Univer. Press, 1959.

ROCK, I. The role of repetition in associative learning. *Amer. J. Psychol.,* 1957, **70,** 186–193.

SPENCE, K. A. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1955.

SPENCE, K. A. Behavior theory and selective learning. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln: Univer. Nebraska Press, 1958.

TOLMAN, E. C. *Purposive behavior in animals and men.* New York: Appleton-Century, 1932.

VOEKS, V. W. Gradual strengthening of S-R connections or increasing number of S-R connections. *J. Psychol.,* 1955, **39,** 289–99.

ON PROBLEMS OF CONDITIONING DISCRIMINATED LEVER–PRESS AVOIDANCE RESPONSES [1]

DONALD R. MEYER, CHUNGSOO CHO, AND ANN F. WESEMANN

Ohio State University [2]

The learning of avoidance, as Solomon and Brush (1956) have remarked, is something that we all take for granted. But, and they have put it particularly well, the

failure of *S*s to learn to avoid are *not rare*, but they are less apt to be reported in scientific papers than are the successes. And the conditions surrounding failures to learn to avoid are as instructive to the investigator, in many cases, as are the optimum conditions for producing the avoidance phenomenon.

It is in this spirit that we offer certain data concerning failures to avoid. These have been collected in both formal and informal studies that have now occupied the better part of some two years. All have been conducted with rats, and all within the format provided by one or another leading brand of commercial lever-pressing apparatus. Some slight differences have been observed between the different brands, but we can conclude at the outset that the problems herein described are not unique to any variant. Rather, the difficulties seem to be inherent in the lever-pressing situation, and not in different constellations as provided by the different fabricators.

THE BASIC PROBLEM

We must emphasize that such equipment *can* be used in training rats to make stimulus-discriminated lever-press avoidance responses. Our concern is wholly with the rate at which such learning is established. We had first obtained this apparatus for some studies of the functions of the visual cortex; these investigations had been planned as extensions of the work of Meyer, Isaac, and Maher (1958). The bar-press situation was selected because we wanted to examine two responses that would be the same except for the incentive under which their learning was established. Further limitations thus imposed were two: that light be the warning stimulus, and that the habit be learnable within a single day. These did not, at first, appear to be unreasonable requirements. The primitive device that we had used before had worked well enough to give us learning in very rarely more than 100 light-shock presentations. It was our presumption that the lever-pressing habit could not be established quite so quickly, but we readily accepted this because of the convenience that the new equipment afforded. We were not, however, prepared in any way for our first empirical encounter, during which a rat failed to show reliable avoidance after *several thousand trials*.

A few such experiences led us to conclude that we were doing something very wrong; it did not seem possible that lever-press conditioning could be as bad as this for the purpose. However, the first of our inquiries to persons familiar with the bar-press situation did not prove to be encouraging; we learned, for example, that it often takes as much as 40 hours to establish stable levels of performance.

Somewhat more salubrious estimates were nearer to 10, but it seemed that not all rats could be counted on even then. This perhaps should not have shocked us, but it did—for we were remembering with sorrow that Mote (1940) had trained his average rat to turn a wheel within 50 light-shock presentations. The figure of merit for the lever-press response is, by this comparison, so poor that only in an automated era could it possibly survive as an indicant of choice.

In retrospect, so far as our brain work was concerned, our course of action at this point should have been to build a Brogden-Culler apparatus and give up lever pressing forthwith. But it is most difficult to turn one's back to a gleaming, thousand-dollar box; then too, we were still convinced that *some* combination of factors would produce efficient learning. Therefore we proceeded to assess the most obvious of these in an informal manner; bar weights, shock strengths, and light intensities were among the first considerations. Throughout these experiments, the general conditions were all very much the same: the warning stimulus consisted of a sharp illumination of the box; the box was darkened when the bar was pressed, and this response terminated shock on trials when the shock was not avoided.

There was never any problem whatsoever in training rats to make escape responses: indeed, after relatively few shaping trials—rarely more than 30 or 40—the latencies on most trials were so short that it could be doubted that the shocks received were noxious. Frequently the rat would learn to hover just above the locus of the bar; this position was maintained throughout the presentation of the warning light. When the shock arrived, the rat depressed the bar very quickly, but in most instances it failed to learn to

do this during the warning interval. To counter this maneuver of "avoidance by escape" we next arranged a program such that shocks could always be avoided, but that periodically were of a fixed duration rather than escapable. This of all the variables informally considered seemed to make at least some difference.

TWO ILLUSTRATIVE EXPERIMENTS

Two formal studies serve to illustrate just how difficult this problem is. We had hoped, because of certain problems of control, to train at least three rats a day; practically, the most that we could possibly invest in training was a 4-hour session. We were therefore driven to consider massing as a possibility and hence trained three groups of nine rats each with intervals between trials of 20, 40, and 60 seconds, respectively. Each trial consisted of light presentation for a 5-second period; pressing of the bar turned off the light and stopped or prevented the occurrence of a shock of 1.6 milliamperes. The outcome, now so typical of many, was that the animals learned virtually nothing during their first training sessions, and this despite the fact that they had first been shaped to make escape responses and then had been presented with from 240 to 720 light-shock pairings.

Some small evidence of avoidance learning did appear when training was continued for a second daily session. The intermittent nonescape procedure then was instituted, and the rats were trained for three more days. But the final outcome, based upon a total of from 1,200 to 3,600 presentations, was that the 20- and 40-second groups still were not avoiding properly. For these groups, in fact, the median percent avoidance scores were .59 and 1.32. The 60-second group was somewhat

better: here at least some animals performed beyond a level of 50% avoidances. Only four, however, did so, and the median performance for the group was mere 6.69%.

At this juncture, we began to speculate that the problem might be in duration of the warning stimulus. This has been manipulated widely in different experimental contexts; Warner (1932), for example, examined a range of from 1 to 30 seconds in a barrier situation. He found that learning was progressively faster as the warning interval decreased, but did not examine the range below 10 seconds in detail. In most relatively recent studies, these have been the intervals of interest, and 5 seconds seemingly is most popular. Mote, on the other hand, presented light for just 2 seconds, and this interval was used successfully by Meyer, Isaac, and Maher. While we could not guess what the optimum would be in lever-pressing situations, it appeared to us that the best place to look was in the range below 5 seconds.

For this experiment, the presentations were given at a rate of one a minute; four groups of rats, 15 in a group, were trained for a total of six hours. There were two daily training sessions, each of which was three hours long; there were thus 180 pairings per day and 360 total pairings. We must say, perhaps, that these are few trials by operant conditioning standards; however, the number is *seven times* as large as that required for learning in a wheel.

The differential treatments were confined to the first training session. Warning stimulus durations prior to shock were, for the four respective groups, 0, .5, 1.5 and 2.5 seconds. During the second session, this was fixed for all at a value of 5 seconds. This was done so that the superiority of one of the differential treatments

might be evidenced without the prejudice of changing latency requirements. Here again, as in all the studies we have done, lever-pressing terminated light; if the animal did not avoid, light and shock went off together.

These procedures yielded median performances approximating 20%; variations for the different intervals proved to be no more than 5%. But even these performances could not be taken as evidence for some avoidance learning; associated with them were bar-pressing rates that often reached extremely high values. There were many instances, in fact, of rates above two thousand per hour; had these rates been evenly distributed in time, at least a third of all the subjects would have shown us ultimately perfect *avoidance*. We obtained, instead, a suppression of the rate when the warning stimulus came on; and this suppression was evidenced throughout the conduct of the study.

INTERPRETATION

Suppression, in itself, is not surprising. It is, indeed, the essence of the readily established CER of Hunt and Brady (1951); we expect it and we almost always find it regardless of our test technique. The question, rather, is why such suppression persists in lever-pressing situations, when its existence presumably implies a drive which lever pressing reduces. If light produces "fear," and shock produces pressing, and pressing eliminates the light, how by any concept of avoidance learning can we fail to get avoidance?

We have found but one asylum, and that is in supposing that the offset of light is *not* an effective "fear-reducer." Having done so we can only wonder why we had expected it to be; in ourselves, we know that "fears" persist,

once evoked, far beyond their instigators. We do not intend this statement as a glove to Mowrer and Lamoreaux (1951) however; all that they have said is that termination "will have the effect of reducing the rat's fear at least a little and will thus reinforce, at least mildly . . ." the response which brings about the termination. Some reinforcement is undoubtedly produced, and we would merely be inclined to substitute the word "trivially" for "mildly."

But why is the particular form of the response so important? Most experimenters with whom we have discussed this problem have considered it to be a matter of the relative availability of the response to be conditioned. However, it is difficult to argue that responses sometimes made as many as 20,000 times are lacking in availability. We believe, instead, that response topography is of consequence for other reasons and that we can best begin to show that this is so by first reviewing Schoenfeld's (1950) concept of the process of avoidance.

While recognizing that a light, such as we have used, becomes aversive after noxious pairings, Schoenfeld has emphasized the role of proprioception as linking avoidance to escape. In a sense, his view is that the rat, in avoiding, is escaping from its own response to the warning stimulus. Termination of the proprioceptive feedback is a reinforcing operation, as is termination of the now-aversive warning stimulus itself. We believe that this approach is basically correct from the standpoint of its emphasis and that we can extend it to interpret the fact that apparatus can be ordered in efficiency for avoidance learning from wheel to shuttlebox to lever.

Briefly, the argument is this. If the termination of a proprioceptive pattern has a reinforcing function, promptness and completeness will determine how effective this will be. Promptness and completeness will themselves depend upon our getting rid of the responses whose presence at the moment has yielded the proprioceptive pattern to be terminated. Some responses will, of course, disappear upon removal of related stimuli, but others will persist unless inhibited through evocation of antagonists. The freezing-crouching pattern, or conditioned "fear," is a member of this latter group; thus the termination of its proprioceptive pattern will require response competition.

Since the freezing-crouching pattern is a posture, suitable antagonists are few; in specifying these, we think we specify the quickly learned avoidance responses. These responses must, in scope, be commensurate with the pattern of conditioned fear, for when this is only partially the case the inhibition cannot be complete. The lever-press response thus is learned but slowly because it is a poor antagonist; its occurrence can be quite compatible with persistence of the prior pattern. But the situation is obviously different with a jump or running response; these yield us competition, termination —and rapid avoidance learning.

Mowrer (1940), twenty years ago, remarked in passing that he had had no luck when he had tried to establish evidence of good avoidance learning with a "type of pedal-pressing response." With his conclusion we must heartily agree, and we think our tests have been exhaustive. However inconvenient the general implication, operants are *not* arbitrary; in avoidance learning, their selection is perhaps the most important of considerations. Possibly, a standard box can be so modified that suitable responses will appear; we ourselves are trying this in a minor way with a "hit-the-

ceiling" pattern. Whether this will work as well as a shuttlebox is something that remains to be seen, but never again will we regard such situations as outmoded and inelegant survivors of an era best forgotten.

REFERENCES

HUNT, H. F., & BRADY, J. V. Some effects of electroconvulsive shock on a conditioned emotional response ("anxiety"). *J. comp. physiol. Psychol.*, 1951, **44**, 88–98.

MEYER, D. R., ISAAC, W., & MAHER, B. The role of stimulation in spontaneous reorganization of visual habits. *J. comp. physiol. Psychol.*, 1958, **51**, 546–548.

MOTE, F. A. Correlations between conditioning and maze learning in the white rat. *J. comp. Psychol.*, 1940, **30**, 197–219.

MOWRER, O. H. Anxiety-reduction and learning. *J. exp. Psychol.*, 1940, **27**, 497–516.

MOWRER, O. H., & LAMOREAUX, R. R. Conditioning and conditionality. *Psychol. Rev.*, 1951, **58**, 196–212.

SCHOENFELD, W. N. An experimental approach to anxiety, escape, and avoidance behavior. In P. J. Hoch & J. Zubin (Eds.), *Anxiety.* New York: Grune & Stratton, 1950. Pp. 70–99.

SOLOMON, R. L., & BRUSH, E. S. Experimentally derived conceptions of anxiety and aversion. In M. R. Jones (Ed.), *Nebraska symposium on motivation, 1956.* Lincoln: Univer. Nebraska Press, 1956. Pp. 212–305.

WARNER, L. H. The association span of the white rat. *J. genet. Psychol.*, 1932, **41**, 57–90.

ON THE TRAINING OF ORIGINALITY [1]

IRVING MALTZMAN

University of California, Los Angeles

There has been a growing research interest in originality during the last decade. Areas of major concern have been the study of the productivity of creative individuals (Dennis, 1958; Lehman, 1953; Taylor, 1956, 1958; Van Zelst & Kerr, 1951), their personality (Barron, 1955; Drevdahl, 1956; Taylor, 1956, 1958), the sociological and cultural factors promoting creativity (Barnett, 1953; Stein, 1953; Taylor, 1956, 1958), and investigations designed to develop reliable paper and pencil tests for the study of individual differences (Guilford, 1950; Springbett, Dark, & Clarke, 1957; Wilson, Guilford, & Christensen, 1953). [2]

Although there is general recognition of the importance of the problem of devising training techniques for facilitating creativity, most of the work in this important area has been anecdotal or merely hortatory (Ghiselin, 1955; Mearns, 1958; Osborn, 1957; Slosson & Downey, 1922).

SCOPE OF THIS PAPER

The purpose of this paper is to review the experimental research that may be relevant to the problem of devising techniques for increasing originality and to indicate the behavioral principles possibly involved in the production of the desired effect.

[2] These are only selected references and are not intended as an exhaustive survey of the relevant literature. A more complete list of references on originality may be found in Taylor (1956, 1958).

For purposes of further discussion we shall distinguish between originality and creativity. Originality, or original thinking, as we shall use the term, refers to behavior which occurs relatively infrequently, is uncommon under given conditions, and is relevant to those conditions. Criteria of relevance and uncommonness, of course, must be established for any given situation. This is a problem more readily solvable within the laboratory than without. Creativity, according to the present usage, refers to products of such behavior and the reactions of other members of a society to those products. An invention is a creative product that may have an important effect upon society and is a consequence of original behavior. But considerably more variables enter into the determination of creative works than originality alone. Our distinction implies that an individual may be highly original but not creative. His brilliant theories are never published, or they are not implemented by the necessary research. Perhaps the original research is conducted, but the results are not published or patented. All this may be done, but the work may be overlooked or the implications disregarded by the appropriate community of scholars or society at large. Further common illustrations may be given, indicating that a great many more behavioral and societal variables influence creativity than originality, making the study of originality under simplified laboratory conditions more feasible than that of creativity. Many of these variables may affect original behavior as well, but they are not of

such overwhelming importance when the training of originality is considered. Our subsequent discussion therefore will be concerned solely with originality as here defined. One further distinction in the experimental study of original behavior needs to be made. Such behavior is always relative, either to a given individual's past behavior or to the norms of a population of which he is a member (Barron, 1955). The two are no doubt correlated. But it is the latter which is related to creative works; and since all of the pertinent experimental research has been of this kind, we shall be concerned only with the study of originality relative to some group norm.

Guilford (1959) also distinguishes between originality and creativity. However, the basis for the distinction stems from his factor analytic approach. Originality, defined essentially in the same fashion as in the present paper, is one of several behavioral traits contributing to creativity. The latter is a more general behavioral trait including as components several forms of flexibility, fluency, and motivational and temperamental traits in addition to originality. Although this is a reasonable empirical classification of traits based upon factor analysis, it does not necessarily contradict the present one. We do not, however, accept Guilford's additional assertion that an S-R approach cannot deal adequately with creativity.

EARLIER ACCOUNTS OF ORIGINALITY TRAINING

Although psychologists have done relatively little experimental research on the training of originality, nonpsychologists have been concerned with the problem for some time. What is more, their suggestions, although not supported by evidence meeting the usual scientific standards,

are frequently in accord with commonly accepted behavioral principles.

For example, Mearns (1958) in a book originally published in 1929 repeatedly emphasizes that in order to facilitate the originality of school children in the arts the teacher must reinforce, manifestly approve, his original efforts. He illustrates many techniques for inducing the child to make public his efforts, to gain the confidence of children so that communication will occur with their "secret unexpressed selves." Furthermore, the teacher is advised to approve of only the genuinely original effort. The teacher is further cautioned to wait patiently for the appearance of original behavior which is fostered by a "permissive atmosphere," the absence of "drill" and excessive discipline. According to Mearns, original behavior appears eventually because all normal children have an urge, energy, or impulse to be creative.

Aside from the romantic approach to a problem of behavior, we would agree that the way to foster originality is to reinforce such behavior when it occurs. A basic difficulty is that it may not occur at all or at such infrequent intervals that the reinforcements cannot shape up such behavior. Thus the fundamental problem in the training of originality is to devise methods for increasing its occurrence in diverse situations, is to get it to occur in the first place thereby permitting the operation of reinforcement. If the operant level of originality is too low, conditioning will not be effective, training cannot occur. We are assuming, then, that originality can be learned and that the same principles of conditioning hold as in other forms of operant behavior. Once it occurs it can be reinforced, approved, recognized. It will then show an increased probability of occurrence. An appar-

ent difficulty, however, is that originality is manifestly different from other behaviors. How can the reinforcement of one bit of uncommon behavior increase the frequency of other uncommon behaviors which by definition are different. Despite the apparent paradox, this sort of non-specific transfer does occur, as we shall see.

Royce (1898), as far as we know, was the first to experimentally attack the problem of devising techniques for facilitating the occurrence of originality. He employed two different procedures. First, Ss were asked to draw a series of nonsense figures unlike anything they had ever seen. A second procedure was to present a series of drawings to S with instructions to draw figures as different as possible from each of the models. Examination of the qualitative changes that occurred under these conditions led Royce to conclude that these procedures may facilitate originality and that they correspond to conditions in society that are conducive to originality.

Slosson and Downey (1922) suggest another method for training and testing originality in writing fiction. They suggest writing a plot, characters, or an entire story, based upon the more unusual messages found in the personal columns of newspapers. Headlines are also proposed as training materials. Variations are possible in the use of the materials by placing restrictions on the kind of plot or characters to be invented. Thus, given an unusual message, the writer is forced to practice literary invention. The assumption is that this form of "mental gymnastics" will increase the originality of his writing.

A simple test of originality suggested by Slosson and Downey (1922) is to determine how many different plots can be written to the same message, and for how many messages can plots be written in a given time interval. A more formalized test is also presented, but no empirical data are given from its use. Likewise, no data are presented on the effectiveness of the training method which appears to have been employed as a teaching technique in creative writing courses at the college level.

Osborn (1957) also believes that originality can be trained and that practice in producing original "ideas" will further develop originality. He gives many exercises designed to provide such training. They typically consist of various kinds of problem situations for which as many different solutions as possible are to be offered. The quantity of different "ideas" is stressed, since it is believed that the likelihood of original "ideas" increases with increases in quantity.

Osborn (1957) has also initiated a procedure of "group ideation" known as "brainstorming" which is designed to evoke original "ideas," and which has received considerable attention from the public press. It is a modified group free association procedure with emphasis upon producing large numbers of "ideas" in the absence of criticism or judgments as to their value. Critical evaluation of "brainstorming" sessions occurs later in order to avoid inhibiting unusual "ideas" during the session.

Although there have been courses established to teach "brainstorming" in business and industrial firms as well as colleges, there is as yet little objective research demonstrating its efficacy.

Taylor, Berry, and Block (1957) found that group brainstorming did not produce a greater number of "ideas" or unique "ideas" as compared to individuals working alone and

whose performance was scored as a nominal group. In fact, it was found that the nominal groups were significantly superior to the brainstorming groups under the conditions of their experiment. Corroborative evidence was found in another study by Taylor and Block (1957). These experiments, however, were not designed to study the effects of originality training in groups or by individuals. They provide no information on the problem of training originality, whether giving many different solutions to problems results in an increased probability of giving original solutions to subsequent different problems.

Meadows and Parnes (1959) have recently reported that a course based upon the principles described in Osborn's book (1957) produced a significant increment in originality in its students as compared to a control class taking a different course. But as in many classroom experiments, there is the difficulty of assessing what the relevant variables are, whether it was the training in problem solving per se, or the changes in motivation resulting from participating in a class and studying a procedure which explicitly claims to facilitate problem solving. There are additional difficulties inherent in the study since no attempt was made to control or assess the extent of differential rehearsal of the test materials in experimental and control classes. The experimental class practiced on materials very similar to the originality tests given before and after the courses were taken. Finally, the design employed permits a biased sampling of Ss, since the "creative problem solving" course was an elective one.

The early psychological literature as well as the suggestions of the nonpsychologists reviewed tend to agree upon a small number of different procedures for increasing originality. One training procedure is to present an uncommon stimulus situation, a situation for which common or conventional responses may not be readily available. Relatively uncommon responses may be evoked as a consequence (Slosson & Downey, 1922; Osborn, 1957). A second procedure is the evocation of different responses to the same stimulus situation (Royce, 1898; Osborn, 1957). Under such conditions the successive responses may become more uncommon. A third training procedure is the evocation of the uncommon responses as textual responses. Several recent experiments have employed this procedure (Judson, Cofer, & Gelfand, 1956; Maltzman, Brooks, Bogartz, & Summers, 1958; Maltzman, Simon, Raskin, & Licht, 1960) and will be discussed shortly. Instructions, as distinguished from training, may also be used to increase originality, and appear to be effective under certain conditions (Christensen, Guilford, & Wilson, 1957; Maltzman, Bogartz, & Breger, 1958; Royce, 1898).

"PRODUCTIVE" THINKING AND ORIGINALITY

Before turning to the current experimental research on originality training, however, note should be taken of several recent studies that have employed Maier's (1931) two-string or analogous problems as a test situation. The reason for commenting upon these studies is the assumption that there is no fundamental difference in the behavioral principles determining originality and problem solving behavior generally. Both involve the evocation of relatively uncommon responses, otherwise the situation would not be called a problem or the behavior original. This is especially apparent in connection with Maier's two-string

problem. Maier (1931) considers the pendulum solution to be a consequence of productive thinking whereas other solutions represent reproductive thinking. As indicated elsewhere (Maltzman, 1955) these problem solving behaviors do not involve fundamentally different laws as demanded by Maier, but only a difference in the extent to which initial response hierarchies are modified. Furthermore, as far as can be determined this also is the basis that Maier used for distinguishing reproductive from productive solutions. The former solutions were given by his Ss upon the initial presentation of the problem. Only after these solutions were given and the Ss were told to find still another solution, which happened to be the pendulum solution, did Maier call it productive thinking. The empirical basis for the distinction between reproductive and productive thinking, then, was the probability of occurrence of the solution under his experimental conditions. The pendulum solution was relatively uncommon and is original in this sense.

Several studies have attempted to determine, with some success, whether verbal responses necessary for the solution of a subsequent problem are present in the initial response hierarchy of Ss (Saugstad, 1955; Saugstad & Raaheim, 1957; Staats, 1957). The determination is made by obtaining word associations to objects later found in the problem situation. Responses necessary for the solution of the problem are presumably uncommon, have a low probability of occurrence in the initial response hierarchy. An association procedure has also been used to determine the uncommonness of responses evoked in the initial response hierarchy and their relation to subsequent success in the two-string problem (Maltzman, Brooks, Bogartz, & Summers, 1958).

A series of experiments by Judson, Cofer, and Gelfand (1956) is a limiting case of the use of textual responses in facilitating "productive" problem solving or originality. They employed the two-string problem as the test and rote learning as the training situation. Instead of attempting to increase the occurrence of uncommon responses during training and determining their transfer effects in the test situation, they induced directly the response chain necessary for a pendulum solution to the problem. A group form of the two-string problem was employed, and the Ss were required to write as many solutions as possible to the problem of grasping both strings simultaneously. The design of the experiment was to prompt a pendulum solution to the problem by presenting an experimental group with different lists of words, one of which contained the relevant response chain "rope," "swing," "pendulum." Additional control groups were employed receiving word lists including none of these words, one or all three but each in a different list. In this study as well as in two replications, the experimental group was superior to the control groups, although the differences were not always statistically significant. A further qualification is that these results hold only for men. No reliable differences were obtained among women in the various conditions. An analogous experiment was conducted employing a group form of Maier's hat rack problem, and comparable results were obtained. These studies suggest that repeated evocation of the critical words during rote learning increased the probability of their occurrence in the problem situation which in turn lead to an increased frequency of the desired problem solution.

Another related study has been reported by Flavell, Cooper, and Lois-

elle (1958). Their study was designed to demonstrate that "functional fixedness" could be reduced by evoking uncommon responses to an object during a pretest period. In one respect it is more pertinent to the problem of producing relatively nonspecific transfer than the studies by Judson, Cofer, and Gelfand (1956), because the responses evoked during the pretest period were not the appropriate ones in the subsequent test problem. However, for our purposes the experiment suffers the limitation that it does not show an actual facilitation of problem solving performance but only a reduction in "functional fixedness."

A series of studies by Maltzman, Brooks, Bogartz, and Summers (1958) combines features of both of these experiments (Judson et al., 1956; Flavell et al., 1958) by evoking uncommon textual responses during a pretest period which are not involved in the problem solution in any obvious manner. Evidence of facilitation was obtained under these conditions. A screwdriver and a block of balsa wood were the objects available for use in the problem situation. Only the screwdriver was sufficiently heavy to serve as a weight permitting a string to swing in pendulum fashion. In the first experiment a control group was given the test problem without any pretraining. Lists of unusual uses for a screwdriver, balsa wood, and string were read by the experimental group before they performed in the problem situation. None of the uses referred to operations involved in the experiment. It was nevertheless found that the experimental group solved the problem significantly faster than the control group, and these results held for both sex groups. Additional experiments with modifications in the lists of unusual uses also obtained reliable differences between experimental

and control Ss, but typically in the case of women and not the men. Many variables in this experimental situation are in need of further analysis. However, the inherent lack of reliability of the two-string problem as a test and the difficulties involved in obtaining experimental control over the relevant variables in the situation suggest that more fruitful studies of originality and problem solving generally should employ another type of experimental situation. Also, with respect to the study of originality, test problems of this sort usually involve uncommonness of response only by inference or by definition. A more direct measure of originality would be highly desirable.

CURRENT STUDIES OF ORIGINALITY TRAINING

Because they seemed to meet the desired requirements, tests of originality of the kind developed by Guilford (1950) and his associates (Wilson, Guilford, & Christensen, 1953) as well as earlier writers (Hargreaves, 1927; Slosson & Downey, 1922) were chosen for a more detailed experimental analysis of variables involved in training originality by Maltzman and his associates (1958, 1960).

In the first experiment in this series (Maltzman, Bogartz, & Breger, 1958) the Ss were run individually. All of the subsequent studies were conducted in group form. A modified free association procedure was employed in which the stimulus words were selected on the basis of their restricted response hierarchies. Since a relatively small number of different responses are evoked by the stimuli, a high degree of communality occurs among their responses. A control group received an initial list of 25 words to which they gave free associations and a different final list of 25

words to which they gave free associations. This was followed by the Unusual Uses Test of originality (Guilford, 1950). The latter test consists of a series of names of six common objects for which S must give different uses other than their common everyday use.

An experimental group received the same treatment except that the initial list was presented five additional times with instructions to give a different response on each repetition. A second experimental group was treated in the same fashion except that E said "good" after every approximately fifth uncommon response during training. Each of these three groups was subdivided prior to the introduction of the free association test list. Half the Ss in each condition were instructed to try to be as original as possible, and half were told nothing. Each of these subgroups were further subdivided prior to administration of the Unusual Uses Test and instructed to be as original as possible. These latter instructions may have been redundant, if not confusing, since the written instructions to the test already implied as much.

Originality measures were obtained for performance on the free association list by determining the frequency with which the responses to each stimulus word occurred. The mean frequency of his associative responses was taken as each S's originality score. Thus the lower the score the more uncommon are S's responses on the average. The total number of unique responses given to the items was taken as the measure of originality on the Unusual Uses Test, where a unique response was defined as a use occurring only once in the sample. Results of the experiment showed that both training and instructions produced a significant increase in originality on the

free association test list. What is more, the experimental subgroups receiving training in addition to instructions were significantly more original than the control group receiving instructions alone, indicating that the training may produce its effects independently of instructions to be original. Results obtained from the Unusual Uses Test were ambiguous, yielding a significant triple order interaction where the experimental subgroup without originality instructions was significantly more original than its control group. However, results on the Unusual Uses Test from subsequent experiments have consistently shown reliable training effects.

Despite the above evidence indicating that the training procedure facilitated originality, additional information is needed to determine precisely what characteristics of the procedure are responsible for the effect. Differences between the experimental and control groups may have been due to the difference in number of stimulus words presented, number of responses evoked, number of different responses evoked, etc. The fact that different responses were evoked by the same stimuli may be neither necessary nor sufficient for an effective training method. Additional research was therefore conducted in an attempt to isolate the relevant variables.

In one study (Maltzman et al., 1960) the control condition and the experimental condition without verbal reinforcement or instructions to be original were again employed. The latter group received instructions to give a different response to each presentation of the same stimulus word. For convenience they will be called the standard control and experimental conditions, respectively. Two additional experimental groups received lists of 125 different stimulus words,

the same number as the repetitions of the initial list for the standard experimental group. One group received high frequency count stimulus words while the other group was presented with relatively low frequency count words (Thorndike & Lorge, 1944). This procedure equated the experimental groups for the total number of different responses. They differed in that the responses all occurred to different stimulus words in the new experimental conditions whereas different responses occurred to the same stimulus words in the standard experimental condition. Another control group was added which received the same number of repetitions of the initial list as did the standard experimental group, but these Ss were instructed to try to give the same response at each repetition of a given stimulus. This condition would aid in determining whether the number of responses per se, or the number of different responses, is the relevant variable influencing test performance.

Results from the free association test of originality showed that the three experimental groups did not differ significantly from each other and the two control groups did not differ significantly. But each of the experimental groups was significantly more original than each of the control groups. On the Unusual Uses Test the standard experimental group did significantly better than all the other groups. The two experimental groups receiving different stimulus words during training did not differ from the standard control group, while the new control condition was significantly less original than each of the other conditions.

The results from this experiment indicate that under some conditions the evocation of a relatively large number of different responses to different stimuli may increase originality to the same

extent as the method of repeated evocation of different responses to the same stimuli. However, the two methods apparently differ in the extent to which they produce transfer effects, as indicated by the significant difference between the two methods on the Unusual Uses Test. These results show that the evocation of different responses to the same stimuli is an important aspect of the standard experimental procedure, although arousal of many different responses by different stimuli is also effective, within limits.

The significant loss of originality by the new control group was not entirely unexpected, and shows that a training technique can significantly decrease as well as increase originality, a result which is of theoretical interest as well as suggestive for classroom practices. Exactly why the effect should appear on the Unusual Uses Test and not on the preceding free association test is not clear, and is a problem that requires further experimental investigation.

Another experiment (Maltzman et al., 1960) was designed to further explore possible relevant variables and alternative procedures for facilitating originality.

The standard control and experimental conditions were again employed. Another experimental condition was given repeated presentation of items from the Unusual Uses Test in order to determine whether the effectiveness of the training procedure is limited to only certain kinds of materials. This was accomplished by presenting the list of six items a total of six times with instructions to give a different use prior to each repetition. A different kind of training procedure was employed with a third experimental group. They were presented with a booklet consisting of 125 different pairs of words. The Ss were

instructed to indicate the member of each pair which goes more readily with the stimulus word orally presented to them. The same stimulus words and number of repetitions as employed with the standard experimental conditions were administered. Each pair of words were unique responses to the stimulus words selected from the previously established norms. A fourth experimental group was given the word pairs and asked to indicate which member of each pair they thought to be more familiar. The stimulus words were not presented. Results of the free association and Unusual Uses Tests showed that the standard experimental condition was significantly more original than the other four groups on both tests, while the other groups did not differ significantly from each other.

In one of the new conditions uncommon responses were evoked as textual responses following the stimulus words. In the other condition only the textual responses occurred. These conditions were introduced in order to determine whether the uncommon responses must be evoked by the stimulus words in order for the training to be effective. The fact that these groups were significantly less original than the standard experimental group suggests that, at least with the training and test materials employed, uncommon responses must be evoked as intraverbal responses by the stimulus words if there is to be an increase in originality in the test situations.

Since the experimental condition which received original training with items from the Unusual Uses Test failed to show a significant training effect, another experiment was conducted to determine whether the failure was due to an inadequate amount of training. Four conditions were employed in this study receiving different amounts of training. None of the groups showed a significant increase in originality on the free association test when compared with the standard control condition.

Results from the latter experiment and one condition from the previous experiment showed that training with unusual uses items in the same fashion as word association materials does not produce a comparable increase in originality. Thus the method of repeated evocation of different responses to the same stimuli does not work with all materials.

A basic assumption of our approach is that, if original behavior can be induced to occur repeatedly through some training procedure, it will increase in its frequency of occurrence in new situations. If this condition is not met, transfer of originality to new situations cannot occur. Inspection of the unusual uses training data indicated that an appreciable increase in originality did not occur during training. In the case of training with free association materials a marked increase does take place.

Another experiment was conducted in which different groups received 0, 1, 5, or 10 repetitions of the initial word association list before testing on another word association list and the unusual uses. A significant trend was evidenced on both originality tests, with a single repetition of the training list producing a significant increase in originality (Maltzman et al., 1960).

On the basis of the current dearth of relevant information, we can only speculate as to why training with unusual uses failed to induce an increase in the uncommonness of the responses. One hypothesis that may be offered is based upon the apparent relatively limited number of different uses available for each item. Given the instructions to write a different use to each repeti-

tion of an item and the inability of Ss to do so with a different operant, they respond with essentially synonymous uses repeatedly, a condition, as we have already seen, not likely to facilitate originality.

If, as we have assumed, originality can be learned according to the principles of operant conditioning, then the effects of originality training should be of the same kind as those ordinarily obtained in experimental studies of learning. One of the gross characteristics of learning is that the behavioral changes produced by reinforcement persist for some time, are not relatively transistory. Increased originality induced by the training method previously described should therefore also persist. An experiment designed to investigate this property of originality training has been conducted, and positive results have been obtained (Maltzman et al., 1960). Using the standard control and experimental conditions of the previous experiments, a delay of approximately one hour was interposed between the training and test situations for a control and an experimental group, and a delay of two days for another control and experimental group. A significant training effect was obtained on the word association and the Unusual Uses Test of originality.

These results show that the effects of originality training persist to a significant degree for at least two days, under the given experimental conditions. We may conclude from these results and that of the previous experiment that the standard experimental procedure for training originality produces some kind of learning.

RELATION OF ORIGINALITY TRAINING TO LEARNING TO LEARN

Originality training studies of the kind reviewed here and learning to

learn studies (Duncan, 1958; Harlow, 1949; Morrisett & Hovland, 1959) have the common characteristic of producing relatively nonspecific transfer effects. Training on given stimuli transfers to other stimuli in the absence of any obvious opportunity for the operation of primary stimulus generalization. Despite the apparent similarity between the results of training in learning to learn and those discussed here, differences in the experimental operations and the kind of behavior induced in these procedures suggest that different principles are involved in these situations.

The learning to learn procedure involves the learning of a great many different discrimination problems. Different stimuli are presented repeatedly until S has learned a differential response to a criterion or has received a given number of trials. The responses are not varied and uncommon responses are not reinforced, nor can they be, because a specified response is correct. Under these conditions Ss show a progressive decrease in the number of trials required to learn new discrimination problems. In originality training, on the other hand, variability of response is induced. Repeated presentation of the same stimuli and reinforcement of a specific predesignated response likewise does not occur in originality training. In learning to learn, the stimuli vary from problem to problem, but the responses do not. In originality training and test situations the stimuli and the responses vary. Thus the two procedures differ in terms of the operations employed to induce nonspecific transfer. Furthermore, the interpretations offered by Duncan (1958) in his comprehensive study of learning sets at the human level do not appear applicable to originality training. The acquisition of observing responses, habits

of looking at details of stimuli, etc. are important in discrimination problems but do not appear to be particularly relevant in originality training situations of the kind described here. Likewise, the analysis of learning to learn by Morrisett and Hovland (1959) in terms of the discrimination between stimulus trace patterns does not appear applicable to originality training.

AN INTERPRETATION OF ORIGINALITY TRAINING

Two basic problems, not unrelated, posed by the experiments on originality training by Maltzman and his associates (1958, 1960) demand an interpretation. First is the nature or source of the reinforcement for originality. In all of the experiments, successive training trials in the standard experimental conditions are accompanied by increases in the uncommonness of the responses. The originality induced in this manner transfers to different stimulus materials, and these effects tend to persist. Originality training thus produces behavioral changes that are characteristic of learning, yet they occur in the absence of differential reinforcement administered by E. Nevertheless, we believe that uncommon responses received differential reinforcement. Reinforcement was obtained, because the occurrence of any previously established intraverbal association is self-reinforcing. As Thorndike puts it: "The mere occurrence of a connection strengthens it" (1949, p. 23). Thus, where there is an initial probability greater than zero that a given verbal stimulus will evoke a verbal response in an operant situation such as word association, the occurrence of that response will increase the probability of its occurrence on a subsequent presentation of that stimulus. Different initial conditions

may only increase or decrease the amount of reinforcement affecting the stimulus-response connection. In the type of experimental situation with which we are concerned here, the uncommonness of the response evoked by the stimulus is one of these conditions. In other words, the amount of reinforcement is inversely related to the initial probability of a verbal stimulus evoking a verbal response. Originality is more reinforcing than commonplace responding.

An impression gained from observing Ss in the experimental situation is that repeated evocation of different responses to the same stimuli becomes quite frustrating: Ss are disturbed by what quickly becomes a surprisingly difficult task. This disturbed behavior indicates that the procedure may not be trivial and does approximate a nonlaboratory situation involving originality or inventiveness, with its frequent concomitant frustration. This impression of frustration accompanying the task of producing repeated uncommon responses would lend support to the notion that when such responses do occur they are self-reinforcing to a considerable degree. These speculations, however, are not essential to the investigation of the initial hypothesis of the self-reinforcing character of uncommon responses. However, these as well as other considerations suggest that the distribution and scheduling of the evocation of uncommon responses may be of great importance in determining the extent to which a transfer effect will be obtained.

It should be noted, however, that the role of self-reinforcement is probably less significant in producing an increase in originality during training, the successive repetitions of the same word list, than in the test situation. Since instructions to the standard experimental group prohibits the reoc-

currence of the same response to a given stimulus word, successive responses to that word will be progressively lower in the related response hierarchy. This is sufficient to account for the increase in the uncommonness of responses with successive repetitions of the same stimulus words. However, we believe that a second variable contributes to the increase in originality during the training session, generalization of the reinforcement effects from the evoked response to still more uncommon responses in the same hierarchy. However, the role of self-reinforcement becomes significant when the transfer of originality to the test situations is considered. In this case the carrying out by S of previously administered instructions to give different responses will not account for the greater originality of the standard experimental condition in the test situations.

This leads to the second problem presented by the experimental data: the fact that the evocation of uncommon responses on one list of stimulus words facilitates the occurrence of original responses on a different list of words, and even more perplexing, on the rather different Unusual Uses Test. The problem is to account for the fact that reinforcement of one instance of original behavior increases the response tendency for other instances of such behavior under different stimulus conditions. A relatively nonspecific kind of transfer of training is induced. Without it, the facilitation of originality would not be possible.

We do not have any simple rigorous explanation of this phenomenon at the present time, but offer the following tentative suggestions. The intraverbal associations possessed by a normal educated adult are enormously extensive and complex. In all likelihood, almost every verbal response is associ-

ated to some extent with every other. Furthermore, the intraverbal associations among common verbal responses are stronger than between common and uncommon responses. Likewise, the intraverbal associations among uncommon responses are relatively stronger than between uncommon and common. The evocation and reinforcement of uncommon responses will, through complex kinds of mediated generalization, therefore increase differentially the probability of occurrence of other uncommon responses even though they do not ordinarily occur to the same stimuli as the original reinforced responses. In a similar fashion, the evocation of common responses would differentially increase the probability of occurrence of other common responses even though they may be members of different response hierarchies. An additional characteristic of the standard experimental procedure used in originality training is that it provides for the inhibition of common responses as well as the facilitation of uncommon responses. Through mediated generalization the effects of inhibition will produce a decrement in the excitatory potential of other common responses. It is this characteristic which may be responsible for the more general transfer effects obtained with the standard experimental procedure than with the procedures that induce uncommon textual responses or employ uncommon stimulus situations for evoking, in turn, uncommon responses. None of these other procedures would involve the inhibition of common responses.

It should be clear that the foregoing hypotheses are speculative and their connection with the data presented is tenuous. Considerable careful experimentation tracing the intraverbal associations present in the experimental situation is needed as well as the de-

velopment of laws of verbal compounding before a completely adequate explanation of the basis for originality training can be given. Whether or not such explanations can be given, the method of training originality described in this paper seems worthy of further study in its own right, and potentially may have considerable practical application. This again is a problem requiring sustained experimental research employing different kinds of test materials.

In conclusion, it should be emphasized that little is to be gained by disputing the usage of the term "originality" as it has been employed here, or whether the current experiments described are investigating the essence of originality. The important problem is the experimental one of determining the variables influencing the occurrence of uncommon responses in relatively simple situations and the functional relationships that obtain in these situations. Finally, the extent to which these effects hold for more complex behavioral situations must be determined. Such work is now in progress in the UCLA laboratory. But to designate only highly complex behavior of a given sort as original to the exclusion of simpler better controlled behaviors seems gratuitous, as well as an obstacle to the progress of research in this vital area.

SUMMARY

The basic problem in the training of originality is to devise a means of increasing the frequency of uncommon behavior. Once it takes place it may receive reinforcement and increase the probability that other original behavior will occur. Earlier attempts to devise training methods for originality were briefly mentioned, as well as related studies of problem solving.

A series of experiments by Maltzman and his associates was reviewed, and a procedure which consistently facilitated originality was described. This procedure involves the repeated presentation of a list of stimulus words in a modified free association situation accompanied by instructions to give a different response to each stimulus. Under these conditions the responses become more uncommon. When presented with new stimulus materials, Ss receiving such training are reliably more original than Ss receiving no training. Tentative suggestions as to the behavioral bases for the training effect were given.

REFERENCES

BARNETT, H. G. Innovation: The basis of cultural change. New York: McGraw-Hill, 1953.

BARRON, F. The disposition towards originality. J. abnorm. soc. Psychol., 1955, 51, 478–485.

CHRISTENSEN, P. R., GUILFORD, J. P., & WILSON, R. C. Relations of creative responses to working time and instructions. J. exp. Psychol., 1957, 53, 82–88.

DENNIS, W. The age decrement in outstanding scientific contributions: Fact or artifact? Amer. Psychologist, 1958, 13, 457–460.

DREVDAHL, J. E. Factors of importance for creativity. J. clin. Psychol., 1956, 12, 21–26.

DUNCAN, C. P. Transfer after training with single versus multiple tasks. J. exp. Psychol., 1958, 55, 63–72.

FLAVELL, J. H., COOPER, A., & LOISELLE, R. H. Effect of the number of preutilization functions on functional fixedness in problem solving. Psychol. Rep., 1958, 4, 343–350.

GHISELIN, B. (Ed.) The creative process. New York: New American Library, 1955.

GUILFORD, J. P. Creativity. Amer. Psychologist, 1950, 5, 444–454.

GUILFORD, J. P. Traits of creativity. In H. H. Anderson (Ed.), Creativity and its cultivation. New York: Harpers, 1959.

HARGREAVES, H. L. The faculty of imagination. Brit. J. Psychol., Monogr., 1927, No. 10.

Harlow, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.

Judson, A. J., Cofer, C. N., & Gelfand, S. Reasoning as an associative process: II. "Direction" in problem solving as a function of prior reinforcement of relevant responses. *Psychol. Rep.*, 1956, **2**, 501–507.

Lehman, H. C. *Age and achievement*. Princeton: Princeton Univer. Press, 1953.

Maier, N. R. F. Reasoning in humans: II. The solution of a problem and its appearance in consciousness. *J. comp. Psychol.*, 1931, **12**, 181–194.

Maltzman, I. Thinking: From a behavioristic point of view. *Psychol. Rev.*, 1955, **62**, 275–286.

Maltzman, I., Bogartz, W., & Breger, L. A procedure for increasing word association originality and its transfer effects. *J. exp. Psychol.*, 1958, **56**, 392–398.

Maltzman, I., Brooks, L. O., Bogartz, W., & Summers, S. S. The facilitation of problem solving by prior exposure to uncommon responses. *J. exp. Psychol.*, 1958, **56**, 399–406.

Maltzman, I., Simon, S., Raskin, D., & Licht, L. Experimental studies in the training of originality. *Psychol. Monogr.*, 1960, **74**(6, Whole No. 493).

Meadow, A., & Parnes, S. J. Evaluation of training in creative problem solving. *J. appl. Psychol.*, 1959, **43**, 189–194.

Mearns, H. *Creative power: The Education of youth in the creative arts.* New York: Dover, 1958.

Morrisett, L. Jr., & Hovland, C. I. A comparison of three varieties of training in human problem solving. *J. exp. Psychol.*, 1959, **58**, 52–55.

Osborn, A. F. *Applied imagination*. New York: Scribner's, 1957.

Royce, J. The psychology of invention. *Psychol. Rev.*, 1898, **5**, 113–144.

Saugstad, P. Problem-solving as dependent on availability of functions. *Brit. J. Psychol.*, 1955, **46**, 191–198.

Saugstad, P., & Raaheim, K. Problem-solving and availability of functions. *Acta Psychol.*, 1957, **13**, 263–278.

Slosson, E. E., & Downey, J. E. *Plots and personalities.* New York: Century, 1922.

Springbett, B. M., Dark, J. B., & Clarke, J. An approach to the measurement of creative thinking. *Canad. J. Psychol.*, 1957, **11**, 9–20.

Staats, A. W., Verbal and instrumental response hierarchies and their relationship to problem-solving. *Amer. J. Psychol.*, 1957, **70**, 442–446.

Stein, M. I. Creativity and culture. *J. Psychol.*, 1953, **36**, 311–322.

Taylor, C. W. (Ed.) *The 1955 University of Utah research conference on the identification of creative scientific talent.* Salt Lake City: Univer. of Utah Press, 1956.

Taylor, C. W. (Ed.) *The second (1957) University of Utah research conference on the identification of creative scientific talent.* Salt Lake City: Univer. Utah Press, 1958.

Taylor, D. W., Berry, P. C., & Block, C. H. *Does group participation when using brainstorming facilitate or inhibit creative thinking?* ONR 150-166, Tech. Rep. 1 1957, Dept. Industr. Admin. and Dept. Psychol., Yale Univer.

Taylor, D. W., & Block, C. H. *Should group or individual work come first on problems requiring creative thinking when equal time is devoted to each?* NR 150-166, Tech. Rep. 1 1957, Dept. of Indust. Admin. and Dept. Psychol., Yale Univer.

Thorndike, E. L. *Selected writings from a connectionist's psychology.* New York: Appleton-Century-Crofts, 1949.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Columbia Univer. Press, 1944.

Van Zelst, R. H., & Kerr, W. A. Some correlations of technical and scientific productivity. *J. abnorm. soc. Psychol.*, 1951, **46**, 470–475.

Wilson, R. C., Guilford, J. P., & Christensen, P. R. The measurement of individual differences in originality. *Psychol. Bull.*, 1953, **50**, 362–370.

PREPARATORY SET VARIABLES RELATED TO CLASSICAL CONDITIONING OF AUTONOMIC RESPONSES [1]

WILLIAM W. GRINGS

University of Southern California

Experiments on conditioning with human subjects (*S*s) are complicated by the fact that response change or "learning" can take place at different levels, sometimes facilitating and other times interferring with the experimenter's (*E*'s) goals. Razran (1955) emphasized this point when he distinguished among five relations of conditioning and perception. Of his categories two are particularly relevant here, representing a distinction between "conditioning without perception" and "conditioning with perception," the difference depending largely upon whether the relation between the unconditioned reaction and the to-be-conditioned reaction is perceived by *S*.

As a specific example consider studies conditioning the GSR or change in heart rate with human *S*s. Over twenty years ago Cook and Harris (1937) demonstrated one trial "verbal conditioning" of the GSR in which the strength of the response was not dependent upon the number of reinforcement trials. They concluded, not only that the GSR is verbally conditioned, but also that such "verbal conditioning is impossible to prevent in conditioning by the usual procedure." In spite of this early result, most GSR "conditioning" studies since that time appear to have been conducted either ignoring verbal-perceptual factors or assuming that they are controlled by instructions designed to mislead the *S*s concerning the nature of the experiment.

When the "perceptual" problem has been attacked directly the favored approach has been to manipulate verbal and attitudinal conditions through changes in instructions to *S*, informing him of stimulus relations or asking him to take a voluntary part in the process. Early studies of this type were summarized by Hilgard and Marquis (1940). Later studies have not emphasized autonomic responses, although Lacey and Smith (1954) and Lacey, Smith, and Green (1955) demonstrated heart rate conditioning differences between "aware" and "unaware" *S*s; and Grings and Kimmel (1959) noted marked instruction effects in GSR discrimination learning. References to the evanescent nature of GSR conditioned responses are common (Grant & Schiller, 1953, Grings & Shmelev, 1959), and the implication is usually made that the GSR is sensitive to transient changes in human *S*s.

The above is ample to suggest that the handling of verbal perceptual set factors in human autonomic conditioning remains a problem of major importance. Data available are limited to demonstrations of the existence of perceptual factors, with very few attempts to define operations for their detection, evaluation, and control. Extension of understanding of the conditioning process with human *S*s may result from exploration of new operations for inferring and manipulating perceptual sets in situations resembling classical conditioning.
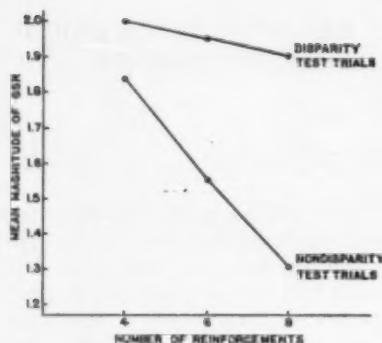
FIG. 1. Average responses on disparity trials and nondisparity trials for groups with different numbers of prior reinforcements. Tone and light stimuli. Response magnitude in square root of conductance change in micromhos.

THE PERCEPTUAL DISPARITY RESPONSE

The present research introduces an operation for inferring the existence of a perceptual set in conditioning studies, one that is not based on verbal instructions to the S. The concept is based on the premise that human Ss, in situations with conditioned stimuli preceding noxious stimulation, use the cue as a preparatory signal—that is, the response to the conditioned stimulus (or what is, in effect, the conditioned response) is partly a preparatory set for receipt of the unconditioned noxious stimulus.

It is proposed that the strength of a perceptual set aspect of conditioned responses can be estimated from perceptual disparity situations, or from situations in which a cue stimulus is followed by an unconditioned stimulus different from that which had been associated with that cue previously. In other words, it is assumed that there will be a difference in magnitude of response between situations where receipt of stimulation is in accord with past experience and where such receipt

of stimulation is not in accord with past experience. It is further assumed that a perceptual disparity variable, so defined, will be related to meaningful determining conditions in human behavior and lead to explanatory possibilities which go beyond more conventional definitions of conditioned responses.

To test the feasibility of such a disparity variable, several experiments have been conducted and will be described briefly below along with later studies amplifying upon the concept. In one experiment, three groups of six Ss each were taught that when they moved a lever in one direction they would receive a tone stimulus and when they moved the same lever in the opposite direction they would receive a bright light. The groups were given different numbers of training trials (four, six, and eight) during which the above regularities applied. Then they were administered a test series containing both disparity and nondisparity trials. Disparity was achieved by following with a tone the direction of movement which previously had been followed by light, or by following with
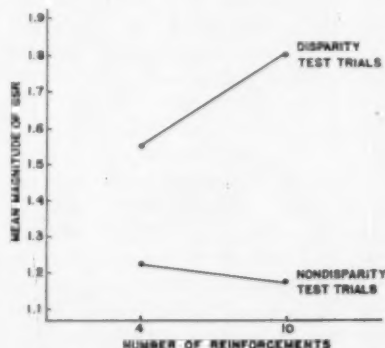

FIG. 2. Average responses on disparity trials and nondisparity trials for groups with different numbers of prior reinforcements. Tone and electric shock stimuli. Response magnitude in square root of conductance change in micromhos.

a light the direction of movement which had previously been followed by a tone.

The results from the test series are plotted in Figure 1 where it can be seen that the magnitude of response on disparity trials was larger than the response on the nondisparity trials for all groups and the difference between the curves increases with larger numbers of training trials. The difference between the disparity and nondisparity responses tested overall by a t test was significant at the 1% level; the effect of the number of reinforcements variable tested as an analysis of variance of the subgroup mean differences between disparity and nondisparity responses was not significant.

A second, similar study (Grings & Kimmel, 1958) was conducted using tone and shock stimuli rather than tone and light as UCSs. Sixty Ss were observed, 30 being given 4 reinforcements before test and 30 being given 10 reinforcements before test. Again the test series contained both disparity and nondisparity trials. The results, shown in Figure 2, agree with those in the first study, i.e., for both groups the response (GSR) on disparity trials is larger than on nondisparity trials (significant at the 1% level) and the size of the difference between the disparity and nondisparity responses increases as a function of the number of reinforcements.

On the basis of these and related studies it was felt that the assumption of a disparity variable was reasonable, and attention was turned to the manipulation of the variable by introducing appropriate independent variation to which it might be related.

Determiners of the Perceptual Disparity Response

On the basis of the already cited studies it was hypothesized that at least two major classes of determiners of the disparity response exist. One involves the strength and stability of the anticipatory response, which in turn appears to be a function of the number of reinforcements of the cue-consequent relations. If preparatory set responses increase and stabilize as most habits do, the more frequent the sequence cue-reinforcing stimulus the more strong and consistent should be the organism's preparation for the reinforcing stimulus, other variables held constant. Elaboration of this point, however, needs to be made, for the operation of competing processes, like response adaptation, is clearly evident in Figures 1 and 2. Further discussion of this matter is made later. In addition, numerous factors related to the conditioned stimulus used (such as its intensity, discriminability, etc.) probably determine the strength of the conditioned preparatory response.

The second major class of determiner of the magnitude of the disparity response would include the direction of the disparity—i.e., whether actual stimulation was stronger or weaker than that for which S was prepared—and the amount of disparity measured on some scale of stimulus distance from anticipated to actual stimulation.

Several studies have been conducted to explore the hypotheses outlined above, two examples of which are given below. One of these has been reported by Kimmel (1960). He was testing the relation between amount of classical conditioning of the GSR and the intensity of the CS. Tone CSs of intensities of 35, 55, 75, 95, and 115 db. re: .0002 dyne/cm^2 were used with an electric shock UCS. In attempting to control differences in response strength to cue stimuli during an extinction series he subdivided each acquisition CS intensity group into three extinction intensity subgroups (45, 75, and 105 db.). Since at the time of the first
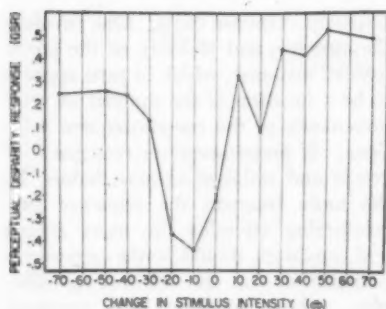
FIG. 3. Magnitude of perceptual disparity response as a function of amount and direction of stimulus change. (Reproduced from Kimmel, 1960.)

extinction trial S had just been given 20 trials during which the first stimulus he received on each trial was the acquisition intensity, it may be assumed that between trials he remained prepared for that intensity. When, therefore, he received an intensity different from the previously experienced one, he might respond to the disparity. Kimmel's data are presented in Figure 3 in terms of GSRs resulting from different amounts and directions of change. The effects of both direction and amount of change were reported as statistically significant.

Similar results were obtained in a study by Grings, Dossett, and Honnard (1959). Forty-eight college students were randomly separated into two experimental groups. One group was given conditioning training with 100% reinforcement, the other with 67% reinforcement. Two cue stimuli (colored light CSs) were followed by a 50-db. tone or a tone of different intensity. Four subgroups of 6 Ss each were trained to different second stimuli of either 30, 70, 90, or 110 db. re: .0002 dyne/cm². A standard number of eight reinforcements of each cue-tone combination was followed by a disparity trial. Since each disparity trial

was made up of the cue to the 50-db. stimulus followed by the actual "other" stimulus, there resulted four different degrees of discrepancy: -20 db., 20 db., 40 db., and 60 db.

The results of this experiment are presented in Figure 4 where mean response to the UCR component of the disparity trial (adjusted for reference level of the UCS involved) is given as a function of amount and direction of disparity. When an analysis of variance was made on the data shown, the difference between the 100% and 67% groups was found to be significant beyond the 5% level ($F = 6.80$, $df = 2$ and 6). The difference between the disparity subgroups, as based on the values portrayed by the combined groups curve in the figure, fell just short of the 5% level of significance ($F = 2.25$, $df = 4$ and 6).

On the basis of the above studies. it is tentatively concluded that the greater the amount of discrepancy between anticipated and actual stimulation the
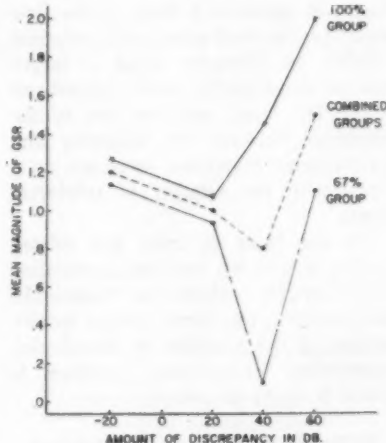

FIG. 4. Mean responses to the UCR component of the disparity trial, adjusted for reference level of the UCS involved, for groups differing in the percentage of prior trials reinforced.
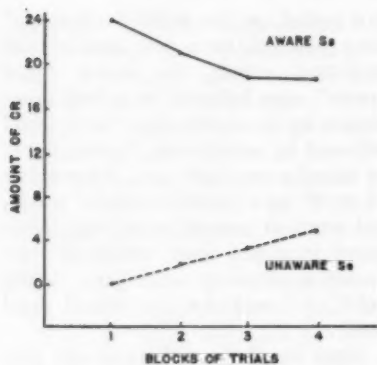
FIG. 5. Comparison of conditioned heart rate responses for groups differing in awareness of CS–UCS relations. (Adapted from Lacey, Smith, & Green, 1955.)

greater the GSR elicited by the actual stimulus. Discrepancies in an upward direction appear to yield larger disparity responses than discrepancies in a downward direction. In addition, these factors appear to be a function of the number of reinforcements and the percentage of reinforcements employed.

APPLICATIONS OF THE DISPARITY CONCEPT

Thus far the emphasis has been upon the definition of the disparity response and the exploration of some of its determining conditions. Next, attention is paid to application of the disparity situation in attempts to clarify some problems of human conditioning.

One of the problems that has concerned the writer is the nature of the GSR acquisition "learning" curve. In other words, what is the effect of the number of trials (or reinforcements) on the magnitude of a conditioned autonomic response? If the conditioned GSR, for example, is determined by interactions of different processes (i.e., nonperceptual and perceptual), it would be predicted that at times one process or the other might be dominant. One

result could be that a "typical" learning curve would not be obtained: that is, GSR magnitudes as a function of number of reinforcements would not be an increase in CR magnitude with an increase in trials.

That different "learning" curves might be expected from different perceptual situations has been demonstrated by Lacey, Smith, and Green (1955). It will be noted from Figure 5 that they found "unaware" Ss gave increasing heart rate CRs with more trials, whereas "aware" Ss showed decreasing CRs with more trials. Similarly, Kimmel (1959), with GSR conditioning in this laboratory, found magnitude of anticipatory CRs to increase up to eight reinforcements, then to decrease. His data are given in Figure 6. The implication in both cases is not that less "learning" is occurring with more reinforcements but rather that the perceptual aspects of the situation are probably changing.

If it might be presumed that perceptual set factors are stabilized by repeated trials where a given CS–UCS relation obtains, it might be proposed that as long as no irregularities occur in the training situation S will cor-
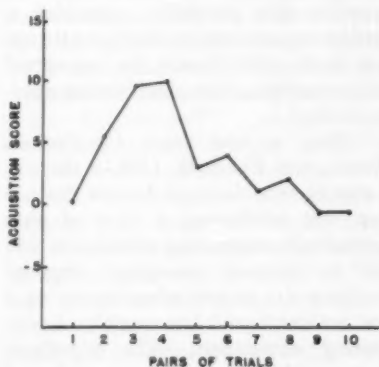

FIG. 6. Acquisition scores for GSR anticipatory CRs in successive blocks of two trials. (Reproduced from Kimmel, 1959.)

rectly "set" for the noxious UCS and therefore will show a curve of adaptation with successive stimulation (just as he would to a series of presentations of the UCS alone). The result will be that anticipatory CRs and UCRs may both show reduced amplitude with increased trials.

It might be argued further that it is the perceptual tendencies or sets which are becoming stronger with increased experience or reinforcements. If the disparity response is taken as evidence of perceptual set strength, it can be predicted from data already at hand (Figures 1 and 2) that the disparity variable (the *difference* between the response magnitudes on disparity and nondisparity trials) does increase with larger numbers of reinforcements. It was also found (Figure 4) that the disparity variable was greater for 100% than for 67% reinforcement conditions.

The proposal here, then, is that the disparity response may show a different course of development (with regard to the number-of-reinforcements variable) than will a simple anticipatory CR. Depending upon conditions it may provide a "better" estimate of strength of preparatory set than the anticipatory CR does. In order to examine this possibility, consider a current experiment in which an attempt was made to emphasize the perceptual set properties of the conditioning environment.

Taking a lead from Cornbecker, Welch, and Fisichelli (1949) the experiment was designed to test the notion that reinforcing a class of cues intrinsically suggesting stimulation will lead to increased perceptual response tendency (to expect stimulation) on a cue not reinforced but capable of suggesting stimulation. The hypothesis was tested by presenting a series of seven-word stimuli which were appropriately reinforced during the acquisi-

tion period, i.e., the words "soft music" were followed by a few bars of soft orchestral music, the words "cool breeze" were followed by a brief stimulation by an electric fan, "loud tone" followed by actual tone, "green light" by actual green light, etc. After either 14 or 49 such "reinforcements" a critical word or neutral word was introduced in a test series which also included a disparity trial (cue "green light" followed by the actual loud tone).

Since in all cases the cue was presented for 5 seconds before the reinforcing (UCS) stimulus was presented, it became possible to trace the course of anticipatory responses following any particular cue. Also, since long and short reinforcement series were employed, the effect of different numbers of reinforcements on both the regular and the disparity trials could be observed. The previous development would lead to the prediction that anticipatory responding (to the cue "loud tone") would become less with more reinforcements, as would also the response to the actual tone, due to adaptation. On the other hand, the disparity response to tone following another ("green light") cue would increase with greater numbers of reinforcements.

The results for this comparison for 20 Ss receiving 14 reinforcements prior to the disparity trial and 20 Ss receiving 49 reinforcements prior to the disparity trial are shown in Figures 7 and 8. The data follow the prediction closely, the anticipatory responses showing significant decrease from short to long reinforcement series ($t = 2.64$, $df = 38$) and the disparity responses showing a significant increase over the same period ($t = 2.23$, $df = 38$).[2]

[2] Essentially the same relations were obtained with changes in digital blood flow as the response variable.

Explanation of the above results has proceeded from familiar behavioral postulates concerning reinforcement and perceptual sets. The concept of reinforcement designated an empirical state of affairs with specified temporal regularities obtaining between classes of stimuli (CSs followed by UCSs). Similarly, the perceptual set concept was related to such general notions as orientative-investigative reactions of Pavlov (1927), anticipatory goal responses of Hull (1930), and the S–R form of Tolman's expectancy (MacCorquodale & Meehl, 1954). It was later pointed out to the writer that experimentation on disparity variables using different descriptive terms had been carried out in the laboratory of Anokhin in Russia. This led to further considerations, as follows.

Anokhin (1955) was interested in a general theory of adaptive behavior which would help to integrate the activities of physiologists and psychologists. One of his central notions is that of "reverse afferentation," a process of reverse signalization or feedback which follows each reflex act "like an echo" and which informs the organism whether the adaptive effect appropriate
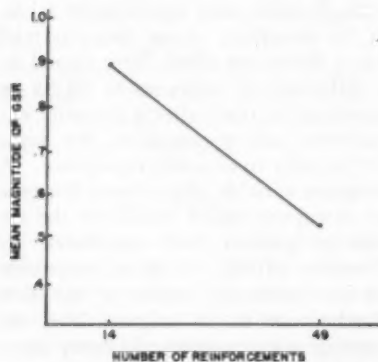
FIG. 7. Anticipatory CRs (magnitude of GSR during delay interval) as a function of the number of reinforced trials.
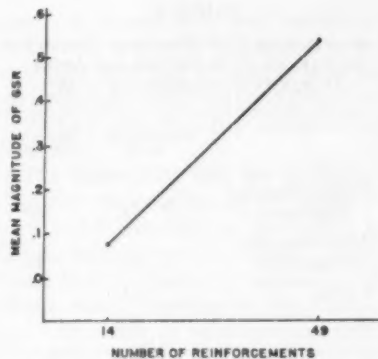
FIG. 8. Magnitudes of disparity response (difference between responses to the UCS on disparity and nondisparity trials) as a function of the number of reinforced trials.

to the situation has been achieved or not. This afferent process provides the necessary conditions to the organism's stopping activity or passing on to other acts comprising the chain of adaptation. The notion of "reinforcement" can be applied to any fractional stage of adaptation which has obtained corresponding reverse afferentation. He talks about "sanctioning afferentation" which stops further attempts at organization of new reflex acts and strengthens the last combination of excitation in the centers of the brain which gave a successful adaptive effect at the periphery.

On the question of "How does the organism tell the sanctioning properties of the afferentation?" Anokhin brings in the concept of "intention." To be effective reverse afferentation must correspond to our intention, that is, it must correspond to some provisional complex of excitations which arose before the reflex itself was formed. At the time of the arrival of the reverse afferent impulses from the UCS to the cortex the behavior will be stable only to the extent that the UCS excitation corresponds accurately to the preparatory afferent excitation which was

TABLE 1

COMPARISON OF GSR RECOVERY QUOTIENTS
TO VARIOUS UCSs BEFORE AND AFTER
DISPARITY EXPERIENCE ($N = 18$)

	Nondisparity UCS	Disparity UCS
Before Disparity Experience	.72	.63
After Disparity Experience	.51	.63
t ratio	2.14	
P	.05	

brought to life by the CS several sec-
onds before the reinforcement.   Put
another way, the afferent apparatus
produces a conclusive evaluation of the
adequacy or inadequacy of the rein-
forcement which follows the signal.
The physiological basis for this process
is conceptualized as a cortical "ac-
ceptor" apparatus.

An experiment cited by Anokhin em-
phasizes the closeness of his work to
that being reported from the USC lab-
oratory.   Two secretory CRs were es-
tablished, one a tone "la" with rein-
forcement with rye rusks on the right
side of the bench and the other a tone
"fa" with reinforcement with rye rusks
on the left side.   During the acquisition
phase no erroneous motor reactions
were made—the animal went right after
"la" and left after "fa."   Then, on a
test day dry meat was placed with the
rusk on the left side.   The result was a
brief investigative reaction and rapid
ingestion of the meat, after which the
previous motor patterns became dis-
rupted.   Regardless of which CS was
given the animal went to the left, sug-
gesting that the eating of the meat led
to a dominance of the reaction to the
left.   Further observation with rusk-
only reinforcements showed refusal to
eat the rusk and marked changes in
the animal's behavior, suggesting what

Anokhin terms a "clear-cut neurotic
state."

The emphasis in Anokhin's experi-
ment would appear to be on the disrup-
tive effect of disparity situations rather
than the basic relation between simple
reinforcement concepts and the concept
of sanctioning reverse afferentation.
However, if one considers the matter
closely, there are numerous implications
for extending the perceptual disparity
research described in this paper   At
least one example will be considered.
From Anokhin's experiment one can
assume that disparity situations pro-
duce marked change in the course of
the "reinforcement" process.   In be-
havioral terms the occurrence of a dis-
parity trial would alert the organism,
lead him to expect a change in rein-
forcement relations.   It could be argued
that even the absence of the UCS, as in
an acquisition test trial, may create
a disparity-like situation.   The result
would be some effect upon the acquisi-
tion process.

Possibilities like the above can be
explored by further analyzing data
from experiments already described.
For example, in the study discussed in
connection with Figure 1, the disparity
test situation contained two series, each
with disparity and nondisparity trials.
If the occurrence of one disparity trial
has a disruptive effect, there should be
a difference in responses to UCSs on
nondisparity trials after a disparity ex-
perience and responses to the same
UCSs prior to disparity experience.   A
response variable which shows this type
of disruptive effect readily is the re-
covery quotient first introduced by
Freeman (1948).   It is an expression
of the percentage return of the skin
conductance to its original level oc-
curring within a standard latency inter-
val.   If the organism's behavior is dis-
rupted, the return to base level should
be slower than when the trials are

proceeding smoothly. Expressed in terms of this experimental situation one would predict a slower recovery from response to the usual UCS after a disparity experience than before. That this is the case can be seen from Table 1.

A similar effect might be expected with magnitude of response measures, that is, a larger GSR would be predicted to a given UCS after a disparity trial than before. In the experiment referred to above this did occur, response magnitude changing from 1.4 micromhos before the disparity trial to 1.7 micromhos after. A related prediction that could be tested readily from numerous published studies is that, if an interspersed acquisition test trial has disparity properties (as it is proposed here it may have), the response to the first UCS after the test trial should be different from the response to the last UCS prior to the test trial. With GSR the magnitudes should be as indicated, i.e., larger UCR after disparity.

These later considerations were introduced only to suggest some of the extended applications the disparity concept may have for explaining certain classical conditioning results. Other relations are being studied currently in order to evaluate the utility of the concept. These extend from attempts to incorporate disparity notions in a rational model of GSR behavior (Dossett & Grings, 1959) to the use of the perceptual disparity response as a measure of individual differences.

SUMMARY

A set of operations for inferring the existence of a preparatory set variable in human autonomic conditioning has been introduced and described. Termed a *perceptual disparity response* it is defined as the difference in magnitude of response between situations where receipt of stimulation following a signal cue is in accord with past experience and where receipt of stimulation is not in accord with past experience with the particular cue. Experimental data demonstrating the disparity variable are presented.

The assumption that the perceptual disparity response would be related to meaningful determining conditions in human behavior was evaluated. It was found that the disparity response magnitude was a function of the number and proportion of reinforcement trials and of the amount and direction of the stimulus change involved.

In an attempt to extend the application of the disparity concept in human conditioning, two further situations were discussed. One concerned the relation between magnitude of disparity response as a function of number of training trials, and compared this to the common learning function relating anticipatory CRs and number of training trials. The second situation evaluated the disruptive effect of a disparity experience in a sequence of regular reinforcement trials.

Brief reference was made to similar concepts and experimental work being carried out at the present time in Russia.

REFERENCES

ANOKHIN, P. K. Characteristics of the afferent apparatus of a conditioned reflex and its importance for psychology. *Voprosy Psikhologii*, 1955, No. 6, 16–38. (Reprinted, Trans. from Russian: The central nervous system and human behavior. U. S. Dept. Health, Educ., & Welfare, 1959. Pp. 241–272.)

COOK, S. W., & HARRIS, R. E. The verbal conditioning of the galvanic skin reflex. *J. exp. Psychol.*, 1937, 21, 202–210.

CORNBECKER, J., WELCH, L., & FISICHELLI, V. Conditioning factors underlying hypnosis. *J. abnorm. soc. Psychol.*, 1949, 44, 359–364.

DOSSETT, W. F., & GRINGS, W. W. A mathematical model for emotional arousal patterns. *Amer. Psychologist*, 1959, 14, 425.

FREEMAN, G. L. *Energetics of behavior.* Ithaca, N. Y.: Cornell Univer. Press, 1948.

GRANT, D. A., & SHILLER, J. J. Generalization of the conditioned galvanic skin response to visual stimuli. *J. exp. Psychol.,* 1953, **46**, 309–313.

GRINGS, W. W., DOSSETT, W. F., & HONNARD, R. R. Conditioned GSR in perceptual disparity situations. *Amer. Psychologist,* 1959, **14**, 393.

GRINGS, W. W., & KIMMEL, H. D. Magnitude of GSR to expected and unexpected stimuli. Paper read at Western Psychological Association, Monterey, California, 1958.

GRINGS, W. W., & KIMMEL, H. D. Compound stimulus transfer for different sense modalities. *Psychol. Rep.,* 1959, **5**, 253–260.

GRINGS, W. W., & SHMELEV, V. Changes in GSR to a single stimulus as a result of training on a compound stimulus. *J. exp. Psychol.,* 1959, **58**, 129–133.

HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940.

HULL, C. L. Knowledge and purpose as habit mechanisms. *Psychol. Rev.,* 1930, **37**, 511–525.

KIMMEL, H. D. Amount of conditioning and intensity of conditioned stimulus. *J. exp. Psychol.,* 1959, **58**, 283–288.

KIMMEL, H. D. The relationship between direction and amount of stimulus change and amount of perceptual disparity response. *J. exp. Psychol.,* 1960, **59**, 68–72.

LACEY, J. I., & SMITH, R. L. Conditioning and generalization of unconscious anxiety. *Science,* 1954, **120**, 1045–1052.

LACEY, J. I., SMITH, R. L., & GREEN, A. Use of conditioned autonomic responses in the study of anxiety. *Psychosom. Med.,* 1955, **17**, 208–217.

MACCORQUODALE, K., & MEEHL, P. E. A reformulation of the theory of E. C. Tolman. In W. K. Estes (Ed.), *Modern Learning Theory.* New York: Appleton-Century-Crofts, 1954. Pp. 177–266.

PAVLOV, I. P. *Conditioned reflexes.* London: Oxford Univer. Press, 1927.

RAZRAN, G. H. S. Conditioning and perception. *Psychol. Rev.,* 1955, **62**, 83–95.

THE PLEASURES OF SENSATION [1]

CARL PFAFFMANN

Brown University

That the senses are the channels of communication between the external environment and the nervous system is so much a truism that, except for those who espouse ESP, the senses are held to be the sole mediators of transactions between the organism and its environment. But students of the senses tend to emphasize primarily the nature of their information handling function, their discriminative capacities and limits as well as the physiological processes upon which they are based. Behavior theorists have given heed to such function in another context when they speak of the cue function of stimuli. According to these conceptions discriminative stimuli serve to steer behavior, to guide it or set the stage for the response. But in this paper I do not plan to elaborate upon such processes. Rather I wish to call attention to certain other aspects of sensory function.

In recent years, neurophysiologists have emphasized that sensory systems may mediate important functions other than discrimination. Attention has been focused on such generalized effects as arousal or alerting of the nervous system as a whole. Bremer (1935) first showed that a state of cerebral somnolence could be produced in experimental animals in which the cerebrum had been surgically isolated from the brainstem and its incoming afferent pathways. He attributed this to the interruption of sensory influx itself. Later analysis (Magoun, 1958) showed that the som-

nolence resulted largely from interference with the centripetal influence exerted by the reticular system, the ascending reticular activating or arousal system. Much subsequent research has shown that this system controls not only the state of arousal of the "higher centers" of the nervous system, but that it can also modulate and influence activity within the classical sensory pathways themselves (Galambos, 1956; Hagbarth & Kerr, 1954; Hernandez-Peon, 1955). The reticular system may also be significant for a variety of other psychological processes (Lindsley, 1951; Schlosberg, 1954).

Most sense organs were found to have at least two central neural pathways. One, the well-known primary projection pathway via the thalamic sensory relays to the cerebral cortex, and the second, a nonspecific pathway by way of the reticular activating system with diffuse projections to the cortex and other neural structures. The reticular system is multimodal and intramodal, for it receives inputs from many modalities. The classical projection systems are specific to one modality and may be said to mediate primarily the cognitive or discriminatory sensory functions, the nonspecific system to mediate physiological and behavioral arousal (Magoun, 1958).

Psychologists especially have emphasized that stimuli may have still other functions, especially that of reinforcement (Hebb, 1958b; Keller & Schoenfeld, 1950; Skinner, 1953; Spence, 1956). This concept has most often been discussed in purely behavioral terms until the more recent studies of direct reinforcement by intracranial

stimulation (Brady, 1958; Miller, 1957; Olds, 1958; Olds & Milner, 1954). In the ensuing discussion I shall limit myself to the so-called *primary reinforcement* for there is almost no limit to the range of previously neutral stimuli that, by one method or another, can be made to acquire reinforcing properties. That *primary aversive reinforcement* has been a function of stimuli has long been known, particularly for those stimuli of high intensity which elicit defense reactions or "reflexes." Certain of these seem qualitatively more prepotent as, for example, the pain of electric shock; "rat runners" have frequently used shock as a primary negative reinforcer in their mazes or lever boxes. Some theorists hold that the pain reduction by shock termination provides the prototype of all reinforcement in the form of drive or tension reduction (Miller, 1959). The study of stimuli as *positive primary reinforcers* has, until recently, been restricted to those stimuli which are naturally related to such biologically functional activities as eating, drinking, or sexual activity. These stimuli could often be assigned a role in processes that mediated the satisfaction of a need or "homeostatic drive." Thus they were said to be related in some way to the primary biological drives. But there has been increasing evidence of late that sensory stimulation, divorced from its need or drive reducing concomitants, may function as a reinforcer in its own right. "Exteroceptive motivation" sometimes called by such names as curiosity or stimulus change (Harlow, 1953; Hebb, 1958a; Kling, Horowitz, & Delhagen, 1956; Montgomery, 1952) has been demonstrated in a number of situations. And one theorist (Hebb, 1958a, pp. 451–467, 1958b) has attempted to link such reinforcement to the reticular activating system on the grounds that changes in level of activation are reinforcing, per se, depending upon the prevailing level at the time of stimulation. Thus, in sensory isolation experiments, "the subject experienced great swings of motivation, which alternated between periods of apathy and intense desire to get back to a normal environment." Hebb concludes, "clearly man's motivation is a function of his exteroceptive stimulation." Thus, although the reinforcing function of stimuli has long been recognized and is the stock in trade of many experimentalists, there is still considerable debate as to the basis for this effect. Does it depend upon some secondary effect as need reduction, drive reduction, or change in arousal patterns in the reticular activating system? Or can we attribute these effects to sensory stimulation directly?

In this paper I shall discuss the proposition that sensory stimulation per se together with its ensuing central neural events be considered as a prime determinant of reinforcement. In this context I shall discuss our experiments on the sense of taste with emphasis on the sensory physiology of taste mediated motivation as revealed in the correlation between gustatory nerve discharges and food preferences in animals. This will lead into a consideration of the relation between hedonic processes and afferent nerve discharges, preference behavior, and taste reinforcement. Finally the relation between affective sensory processes and discriminative functions will be discussed in the light of some speculations on the physiological bases for these two aspects of sensory function and their significance in behavior.

TASTE AS A MODEL SYSTEM

Although I began my studies of the sense of taste with a traditional sensory emphasis, it became increasingly apparent that the gustatory sense has certain unique features for the further
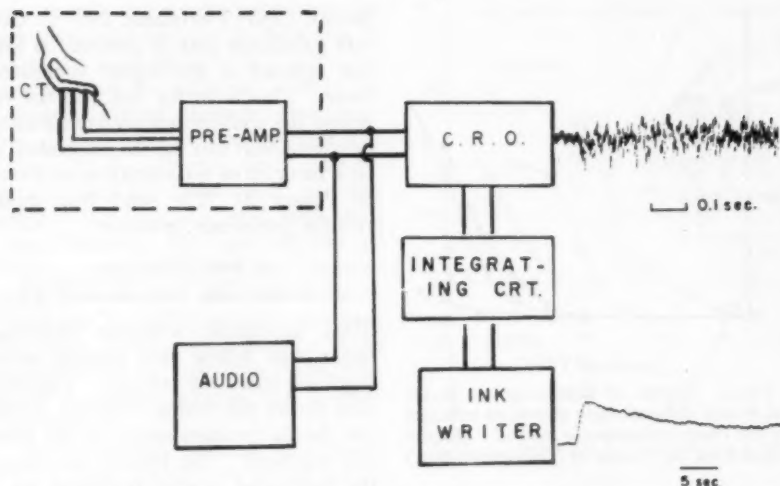
FIG. 1. A block diagram of the recording apparatus showing two types of record. The upper trace shows a typical asynchronous, multifiber discharge from a large number of nerve fibers; the lower trace shows how such activity appears when processed through the integrator. (Reproduced from the *American Journal of Clinical Nutrition*.)

understanding of behavior in general. Not only does the sense of taste possess selective receptor sensitivity that permits the discrimination of different taste stimuli, but these stimuli will elicit a number of specific consummatory responses: drinking and eating or, on the other hand, rejection. It is a relatively simple matter to demonstrate the control of behavior that taste stimuli can manifest by giving the animal a choice of water and taste solutions as Richter (1942) showed in his classic studies of self-selection. In addition, these stimuli can act as powerful reinforcers of instrumental responses leading to ingestion. Thus one and the same sense modality possesses easily demonstrable discriminative and reinforcing functions. In this respect it is an ideal model sensory-behavioral system in which to examine the relation between sensory stimulation and the reinforcing mechanisms in behavior as well as to permit the manipulation of the state of the organism by certain deprivation operations. I should, therefore, like to review briefly some of the methods we have employed and some of our experiments on taste,[2] to show how we look at the sensory continuum and its afferent neural input to the CNS which both motivates and directs behavior. In so doing I shall make use of the behavioral data of other workers, as well as the results of our own physiological and behavioral experiments. Our behavioral studies to date have utilized primarily the preferential ingestion method.

The Gustatory Afferent Discharge and Preference

Afferent nerves, like all nerves of the body, carry a series of electrical pulses which are the signs of impulse traffic up the sensory nerve. These can be
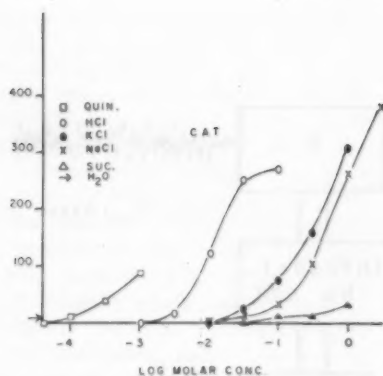
FIG. 2. Curves of taste responses in the cat to four different taste stimuli as indicated by the integrated response method. (Reproduced from the *Journal of Neurophysiology*.)

recorded by appropriate electronic devices as shown in Figure 1 (cf. Pfaffmann, 1959). I shall present only data for the over-all activity in the taste nerve as afforded by the integrating circuit illustrated in the lower trace in Figure 1. The gustatory response' curves for the cat are shown in Figure 2 in terms of the magnitude of the electrical signal generated in the nerve when each of the basic taste stimuli is applied to the tongue receptor surface. By such curves we can map out the taste sensitivity of different animals (Pfaffmann, 1955). These curves do not tell us about "qualitative" differences in the nerve response. For that, analysis of the activity of single nerve fibers in the nerve is required (Pfaffmann, 1959).

From the point of view of behavior, the basic response to taste solutions is either one of acceptance or rejection so that the classical manifold of four tastes, salt, sour, bitter, and sweet, may be reduced to two behavioral classes': acceptance and rejection. We have used a typical two-bottle Richter-type preference situation to study such behavior in different species (Bare, 1949; Car-

penter, 1956; Pfaffmann, 1957). Certain substances may be accepted at low, but rejected at the higher concentrations. This is so for NaCl in the rat where the well-known preference-aversion response can be demonstrated to be a function of concentration as shown in Figure 3. The solid line shows relative preference behavior

$$\frac{\text{cc. intake taste sol.}}{\text{cc. intake taste} + \text{cc. intake } H_2O}$$

Most acceptable solutions including sugar will follow this pattern when ingestion measures are used. Figure 3 also shows the neural response in the rat chorda tympani nerve to the same salt solutions. The broken line shows the integrated neural responses as a percentage of the maximum response observed. Clearly the preference behavior (i.e., response greater than 50%) is not apparent until some substantial neural input has been achieved (20%–30% of maximal discharge). The peak preference occurs at about 60% of maximal input, and the aversion rapidly sets in while the afferent discharge is still climbing. There is no change in the afferent input signal that
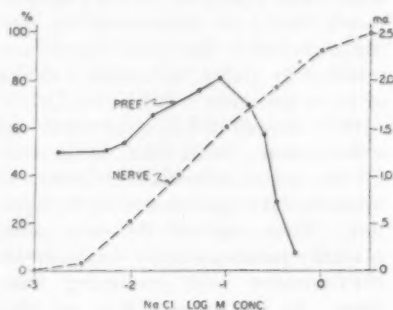

FIG. 3. Preference response to different salt concentrations and magnitude of the chorda tympani nerve discharge in the rat. Ordinate to the left corresponds to preference response, ordinate to the right shows magnitude of integrator deflection. (Reproduced from Pfaffmann, 1958.)

corresponds to the behavioral inversion point—at least we have not been able to detect any such change. Up to the inversion point, the behavior parallels the afferent discharge curve as if the positive approach to the stimulus were due directly to the afferent input. The inversion and the subsequent aversion suggest the intrusion of a secondary "stop" system. The fact that the inversion point is close to the isotonic point for sodium chloride solutions (.9% NaCl [.15M] is isosmotic with the body fluids) suggests that some postingestion or metabolic factor related to water balance has become operative.

Indeed Stellar, Hyman, and Samet (1954) have shown that the salinity of the gastric contents does affect the salt preference. Strong salt solutions intubated directly into the stomach depress the drinking of the more concentrated salt solutions. On the other hand, this is not the only stop factor for, in the same study, animals with an esophageal fistula showed the preference for hypotonic solutions and aversions for hypertonic ones even though the solutions never reached the stomach. Taste factors alone appear capable of eliciting the typical salt preference–aversion function. Thus the stop may indicate a change in "sign" of the afferent input merely as a consequence of its increasing intensity. Perhaps there is a central-neural switching when the intensity of the afferent salt discharge reaches a critical value. Most intense sensory stimuli have aversive effects.

The relation of behavior to sensory afferent discharge for some of the other basic taste stimuli is shown in Figure 4. The upper figure shows the preference response curves for sugar, acid, and quinine as well as salt. We show here only the intakes above 50% for salt and sugar, remembering that at higher con-
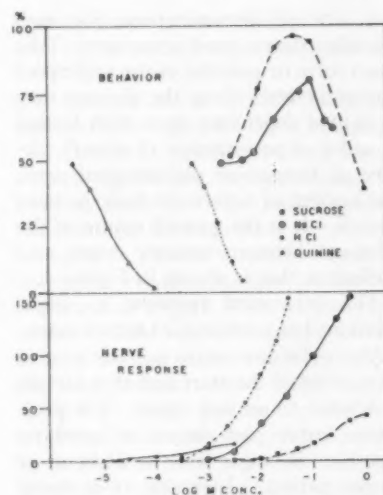

FIG. 4. Composite graph of behavioral and electrophysiological responses in the rat. Upper graphs show the percentage preference (or aversion) as a function of stimulus concentration. The integrator responses in the chorda tympani nerve to the same stimuli are shown in arbitrary units in the lower graph.

centrations, the curves fall below the 50% value. The responses for acid and quinine are both aversions; no preference is shown. Note in the lower curves that the absolute magnitudes of neural response are quite different for the electrolytes as compared with the nonelectrolytes. The behavioral response to quinine is quite definite and appears before there is a clear signal "greater than the noise" for the nerve response. Here the behavioral indicator is the more sensitive. In an analogous way, the response to sucrose behaviorally is clear and definite, yet the neural response is disappointingly small. Intrinsic differences in the magnitude of the neural response, due in part to fiber size, etc., mean that sheer size of electrical signal itself cannot be correlated with behavioral effect. On the other hand, the range of effective stimulus concentration values in the

two sets of determinations for each stimulus shows good agreement. The exact form or position of the preference aversion curves along the abscissa may be shifted depending upon such factors as order of presentation of stimuli, degree of hunger or physiological need, and amount of experience with the taste stimuli. It is the general nature of the relations between sensory input and preference that is shown in Figure 4.

The behavioral response to sugar solutions has been studied rather extensively, and it now seems possible to give an account of the start and stop factors in relation to sensory input. The preference curve just shown is based on ingestion of sugar over a 24-hour or longer period. Ingestion is a rising function of concentration to a peak preference when the typical inflection oc-
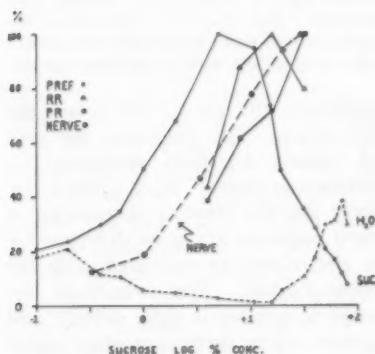
Fig. 5. Composite graph showing the neural response to sucrose and the behavioral measures obtained by three different procedures in the rat. Pref. is the typical two-bottle 24-hour ingestion preference method. Both the sucrose and water intakes are shown as smoothed curves, based on data by Richter and Campbell. RR is the rate of bar pressing during a test period for a small drop of sucrose on a regular reinforcement schedule, and PR is the same on a periodic reinforcement schedule, both based on Guttman (1953). Ordinates have been adjusted to make the maximum values in the original curves equal to 100.

curs and intake declines. This relation is found also in the single stimulus method. McCleary (1953) was one of the first to show that the intake of sugar solution is limited by an osmotic postingestion factor. Others have shown similar effects of gastric factors by studying the effects of intubation upon intake (LeMagnen, 1955; Miller, 1957; Smith & Duffy, 1957; Stellar, Hyman, & Samet, 1954). The relative frequency of choice in the brief exposure preference method, however, does not show the inflection (Young & Greene, 1953). Here there appears to be a linear relation between the level of acceptability and logarithm of concentration. Guttman's (1953) study of the rate of bar pressing as a function of concentration of sucrose solution used as the reinforcement showed that, on a continuous schedule, rate increased for the weaker concentrations but showed an inversion at the higher value. On a periodic schedule, however, the rate of bar press was found to be linear with the log of the concentration. The latter schedule provides relatively little drinking per response. But bar press rate also can be depressed by intragastric injections (Smith & Duffy, 1955).

The relations between bar pressing on the two schedules and the preference ingestion data of Richter and Campbell (1940) are shown in Figure 5 together with the electrophysiological response curve for the rat (Hagstrom & Pfaffmann, 1959). Note that the inversion point in the two-bottle preference test is close to the inversion point of the continuous reinforcement schedule. Both points lie close to the top of the electrophysiological sensory function. On the periodic schedule, rate rises as the sensory function increases. The fact that there is no inversion where the amount ingested is small appears to implicate the postingestion factor as a primary stop mechanism (Collier & Siskel,
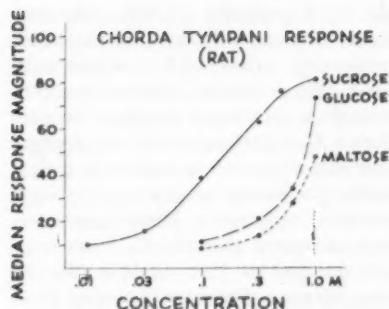
FIG. 6. Integrator response magnitudes to three sugars for the rat chorda tympani nerve. Ordinate is relative magnitude, 100 equals the response to 0.01 M NaCl. (Reproduced from *Journal of Comparative and Physiological Psychology*.)

1959; Shuford, 1959). Such a formulation would agree with Young's (1959) statement "that sugar solutions are basically hedonically positive for the rat as compared with salt which is first positive but then negative."

Studies which compared the effect of two different sugars, glucose and sucrose, on consummatory behavior, choice response, and rate of bar press likewise point to the importance of postingestive factors. Isohedonic concentrations, equally accepted in short comparative tests, were not consumed in equal amount when presented singly for longer periods. The glucose solutions were consumed in less volume than their isohedonically matched sucrose solutions, and the cumulative mean intake of both sugars was linearly related to osmotic pressure for the longer periods (Shuford, 1959). In the Skinner box the concentration of these two sugars which give equal rate of response on an aperiodic schedule (i.e., have equal reinforcing value) corresponds to the isohedonic concentrations (Shuford, 1959; Young, 1957). Guttman (1954) previously showed that the equal reinforcing solutions correspond to the

equally sweet concentrations found in psychophysical experiments (Cameron, 1947; MacLeod, 1952).

Hagstrom and I (1959) have also compared the relative efficacy of sucrose and glucose as gustatory stimuli for the rat, using the electrophysiological recording method. This could be done even though the magnitude of response to sugar is relatively small. Figure 6 shows the relative responses to the two sugars, and Figure 7 shows the comparison of the equi-effective concentrations of the two sugars as determined in the bar pressing experiment, in the preference choice test, and by the electrophysiological measure. All measures agree that sucrose is more effective than glucose of equimolar concentration, although the quantitative relation is not precise, for the electrophysiological data is curvilinear, the behavioral data, linear. However, it should be remembered that the physiological data reflect only the chorda tympani response and not the taste receptors at the back of the mouth. Perhaps the curvilinear relation between the behavioral and physiological measures arises from this.

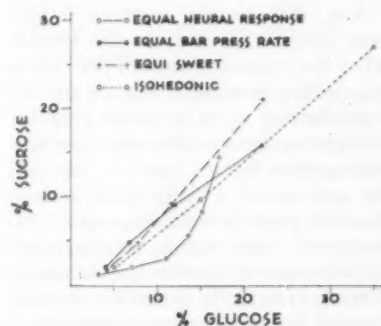Where ingestion is minimal, behavior

FIG. 7. Equal response concentrations of glucose and sucrose for electrophysiological measures, equal bar press rates, equi-sweet solutions and isohedonic solutions. (Reproduced from *Journal of Comparative and Physiological Psychology*.)

and sensory effectiveness seem to go together. Actually the postingestion factor seems to be especially significant in preferences based on ingestion. The glucose preference threshold is lower than that for sucrose and the intake of glucose greater than that predicted by the sensory measures. Soulairac (1947) has shown that the relative preference for sugars in a typical two-bottle preference test is correlated with their relative rates of absorption from the gastrointestinal tract.

Experiments utilizing other learning situations have been concerned with the action of sugar or saccharin as reinforcers (Collier & Siskel, 1959; Hughes, 1957; Sheffield & Roby, 1950; Sheffield, Roby, & Campbell, 1954; Smith & Duffy, 1957; Young & Shuford, 1954, 1955). There is ample evidence that the sweet taste is rewarding whether it is nutritive or non-nutritive. Performance here too is an increasing function of stimulus concentration where conditions appear to maximize the sensory effects and minimize the postingestion factor. When amount or temporal factor is such that the absolute amount or volume of stimulus or nutrient ingested per unit time is low, then purely stimulus factors seem best to account for the results. When the amount consumed per unit is large, either because of a large amount of reinforcing agent or short time between presentations, then the secondary postingestive factors intrude. As Collier and Siskel (1959) point out, it should be possible with the proper combination of concentration, volume, interval between reinforcement, and number of reinforcements to obtain a monotonic stimulus function, a nonmonotonic function with inversion points at different volumes or concentrations, or even no relation at all between amount of reinforcement and performance.

These considerations presumably ap-ply to Campbell's (1958) measurements of the JNR, the just noticeable reinforcing difference for sucrose solutions. He determined the minimal concentration increment necessary to produce a 75% preference for the stronger of a pair of sucrose solutions in a two-bottle preference situation. Stronger solutions required a larger concentration increment than did the weak ones, and the plot of JNR against stimulus concentration showed a U-shaped function reminiscent of the more familiar Weber intensity discrimination functions. The resemblance may be due only in part to the properties of the sensory input because, as we have just shown, preference is the composite of sensory as well as postingestion factors. The osmotic effects of the stronger sugar solutions might be particularly strong. This situation contrasts with the JNR functions for the noxious stimuli where the stimulus itself may be more directly the source of the reinforcement uncomplicated by secondary factors (Campbell, 1958).

Thus we see that sensory stimulation of the mouth receptors, especially of taste, has a direct relevance for the control of ingestive behavior and the instrumental responses which lead to ingestion. Although a number of investigators have provided evidence that the mouth receptors may be bypassed and that learning can still take place, all such evidence shows that such learning is not as effective as when stimulation of the mouth receptors is included (Coppock & Chambers, 1954); Miller, 1957). Nearly all workers, whatever their theoretical predilections, have shown the importance of stimulation and "sensory contact," some would say "sensory satisfaction" (Smith & Duffy, 1957) in the reinforcement process. Indeed as Bindra (1959) concludes in his recent monograph on motivation: "Whatever the interpretation, it seems

clear that, up to a certain point, an increase in sensory stimulation is a positive reinforcer" (p. 134).

The fact that certain taste stimuli control ingestion directly appears to be biologically determined. Frings (1946) has pointed out that nearly all organisms accept sugar solutions. There are exceptions perhaps related to certain aspects of metabolic or other biochemical divergences among the species. Cats cannot taste or discriminate sugar solutions (Carpenter, 1956; Pfaffmann, 1955), and birds (Pick & Kare, 1959) do not appear to show strong sucrose preferences; but these examples are remarkable largely for their divergence from what otherwise appears to be a general rule. Further there is no convincing evidence that the "sweet tooth," where it does exist, depends upon the concomitant nourishment. The drinking of nonnutritive saccharin solutions under prolonged exposure to them shows no sign of extinction such as might be expected if the preference for saccharin or "sweet" were acquired by past association with nourishment (Sheffield & Roby, 1950). Here we have "sweet for sweet's sake."

On the other side of the coin, we have been able to show that a "bitter" aversive stimulus can be made more acceptable *only temporarily* when it is paired with the alleviation of thirst in early infancy (Warren & Pfaffmann, 1958). Newborn guinea pigs were raised on a normally avoided solution of sucrose octaacetate (SOA) as the only source of water for a three-week period. At an older age these organisms showed the usual rejection of SOA. Gustatory stimuli, therefore, appear to be biologically determined as the instigator of consummatory or avoidance responses and as primary positive or negative reinforcing stimuli.

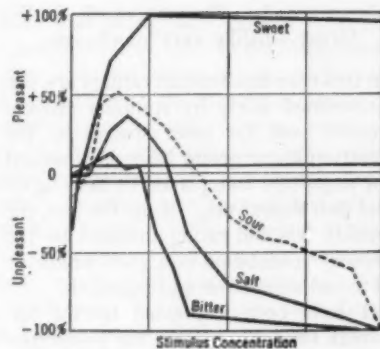So far I have discussed the mechanisms of a variety of behaviors which

FIG. 8. The preponderance of "pleasant" or "unpleasant" judgments in relation to the concentration of taste solution. Ordinate gives percentage "pleasant" minus percentage "unpleasant." The abscissa is proportional to concentration, the full length of the base line standing for 40% cane sugar, 1.12% tartaric acid, 10% NaCl, and 0.004% quinine sulphate (by weight). (Reproduced from Pfaffmann, 1958.)

appear to be under the control of sensory stimulation qua "sensory stimulation." I have not discussed the question of the affective or hedonic aspects of sensory stimulation, the pleasure of sensation, so to speak. To do so, I should now like to turn to other kinds of data derived from studies of man.

Hedonic Aspects of Sensory Stimulation

Of all the applied psychophysical fields, none have made greater use of the affective or hedonic rating scale methods, along with purely sensory testing procedures, than has the field of flavor technology. In a series of carefully controlled tests the Army Quartermaster Food Acceptance Laboratory (Pilgram, 1957), using a nine-point hedonic rating scale, has been able to predict, with good reliability, the actual choices of food and the acceptance of menus on the part of soldiers in the field. These ratings frankly ask such questions as the following:

like extremely—like very much—indifferent—dislike very much—etc.

In this case the hedonic ratings are not determined solely by stimulus properties of food, for other studies by the Quartermaster group have documented the important role played by familiarity and past experience. None the less, the frankly hedonic rating initiated by the sensory stimulation is a good predictor of actual acceptance and ingestion.

I have been impressed by the apparent similarity of our rat preference curves with those of hedonic value obtained many years ago by Engel (1928). In those experiments, subjects were asked to rate different intensities of taste stimuli as either pleasant, unpleasant, or indifferent. The data can be treated in a number of ways, but in Figure 8 we see the ratings of four different taste modalities expressed as percentage of pleasant ratings minus the percentage of unpleasant ratings for a group of seven observers using summed ratings of all observers. The abscissa is proportional to the concentrations adjusted for each of the different stimulus solutions. Note that sugar begins at a slightly unpleasant value and rises with concentration to reach a plateau. Sour, bitter, and salt all start from indiffer-

ence, rise to a peak, and then fall off to unpleasantness. Sweet is predominantly pleasant, bitter predominantly unpleasant, with salt and sour intermediate. In Figure 9 I have plotted the animal preference curve and the hedonic rating by Engel's Ss for sodium chloride solutions by concentration. The hedonic ratings here were computed simply as the percentage of the total ratings that are pleasant. Except for position along the abscissa, the two curves show a striking similarity. Beebe-Center (1951) earlier called attention to these same relations and compared the animal preference responses with hedonic ratings. Actually the more recent analysis of the postingestive factors that control the intake of sucrose solutions has done much to clarify the discrepancy between Engel's hedonic curve for sugar and the rat's preference for sugar. The sucrose hedonic curve does not turn down at the higher concentrations (see Figure 8). As noted earlier, rate of bar pressing on a periodic schedule and frequency of choice in the brief exposure test are likewise monotonically related to concentration. Thus hedonic rating and reinforcement bear the same relation to stimulus concentration.

In the preference–aversion curves for other substances, the stop mechanism may be sensory (gustatory) in origin or might arise from other than postingestive effects. If the taste of strong salt solutions is aversive per se, then we might expect that, regardless of reinforcement schedule, response will fall off above the optimal salt concentration in a manner resembling the "falloff" in Engel's curves. The same might hold for saccharin which has a "bitter" sensory component at the higher concentrations. Further study of these effects is needed.

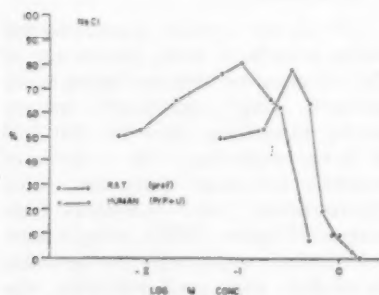I have been emphasizing stimulus properties as prime determinants of the

Fig. 9. Comparison of animal preference and hedonic ratings of man in response to sodium chloride solutions. (Reproduced from Pfaffmann, 1958.)

hedonic effect, but let me hastily point out that specific training and experience may make some unpleasant odors and tastes acceptable or even preferred. I am reminded here of the whiskey drinker's development so aptly described by Brown (1955).

Straight whiskey, when first ingested, typically effects rather violent defense reactions. Because of this, the novice drinker usually begins with sweet liqueurs, "pink ladies," and wines, and slowly works his way through a series of beverages characterized by the gradual disappearance of cola and ginger ale additives. Finally, only plain water or even nothing need be mixed with the raw product. To the hardened drinker, straight whiskey does not taste bad—*not bad at all!* (It is thus that a product euphemistically labelled "neutral spirits" becomes indeed psychologically neutral).

Although this description appeared in defense of a drive reduction formulation of behavior and learning, the mere fact that drive reduction was supported and buttressed by the use of sweet and "pleasant" stimuli seems to have been overlooked. There is no convincing evidence that the primary relation between hedonic tone and sensory stimulation is entirely acquired. Indeed, quite the converse seems to be true.

The relation of sensory stimulation to hedonic process has been noted by earlier workers. Sherrington (1906) noted that the stimulation of contact receptors, as contrasted with distance receptors, is characterized by strong affective tone and that the contact receptors stand in very close relation to consummatory responses. Stimulation of the touch receptors of the lips and touch and taste in the mouth initiates reflex movements that precede the act of swallowing. Troland (1928) spoke of three classes of stimulus reception: nociception, beneception, and neutroception. The principal nociceptor system was pain, but other examples were hunger, thirst, the taste of bitter, strong salts and acids, and certain foul or repugnant odors, etc. In the beneceptor class were the sense organs mediating erotic behavior, the taste of sugar, and certain other odors. Vision, hearing, touch, and proprioception were relegated to neutroception. Although Troland's criterion was that of biological utility, Young (1936) pointed to the high correlation between pleasantness and beneception, unpleasantness and nociception. He also notes that there is a frequent correlation between pleasantness and approach, and unpleasantness and avoidance; and further that the affective responses to simple colors and tones are much weaker than those evoked by odors, tastes, cutaneous or organic stimuli.

In his work on palatability, Young espouses a clearly hedonistic theory of reinforcement in which he treats the affective process as a postulate, an intervening variable (Young, 1959). The reactions of animals to taste solutions then can only be described when reference is made to their positive or negative hedonic effects. Thus, sugar solutions are hedonically positive, and the level of acceptability is directly proportional to the logarithm of the concentration. Salt on the other hand does not show this one-to-one correspondence with sensory intensity. Low concentrations of salt are said to be hedonically positive for the rat, but at high concentrations are hedonically negative. Such terminology often appears simply to rename approach or avoidance behavior with hedonistic synonyms for we all know what we mean when we say in laboratory jargon that the rat likes sugar. We know he will take it in preference to water, he may press a bar for it, he may take it even when it is adulterated with something he "doesn't like" such as quinine. But such usage does not necessarily indicate the primacy of the hedonic process; we must
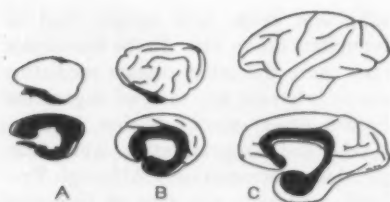
FIG. 10. Lateral and medial surfaces of brains of rabbit (A), cat (B), and monkey (C) drawn roughly to scale. The drawings illustrate that the limbic lobe, represented in black, forms a common denominator in the brains of all mammals. (Reproduced from *Journal of Neurosurgery.*)

have further experimental study of acceptance and rejection and of their relation to hedonic or affective processes.

Thus it is abundantly clear that, instigation of consummatory response (or of rejection), reinforcement of instrumental responses, and elicitation of hedonic effect are all closely related and that reproducible stimulus functions can be demonstrated for each. A definitive choice as to which of these is primary cannot yet be made. I would like to propose that sensory stimulation per se together with its ensuing central neural events be considered as a prime determinant in the chain of events culminating in acceptance behavior, reinforcement, and hedonic effect. The further study of such stimulus functions and their analysis particularly at the physiological level is a problem that should merit the highest priority.

I should like to turn to the possible physiological mechanisms in the reinforcing and hedonic functions of stimuli.

THE PHYSIOLOGY OF REINFORCEMENT AND SENSORY AFFECT

The studies of cranial self-stimulation (Brady, 1958; Olds, 1958; Olds & Milner, 1954) show that the reward and punishment systems within the brain itself, but particularly the reward systems, have been found to depend in

large measure upon the limbic system. In classical neuroanatomy this was known as the rhinencephalon or smell brain, but its relation to olfaction seems less and less particular. The limbic lobe including the hippocampus completely surrounds the hilus of the hemisphere. Its various subcortical cell stations include the amygdala, septal nuclei, the hypothalamus, anterior thalamic nuclei, etc. The limbic lobe of the rabbit, cat, and monkey is shown in black in Figure 10. The limbic lobe appears to form a common denominator in the brains of all mammals.

Its constancy of gross and microscopic structure throughout the phylogeny of the mammal contrasts strikingly with the mushrooming neopallium which surrounds it. . . . Papez theorized that the experimental evidence also points to the limbic lobe as a cortical common denominator for a variety of emotional and viscerosomatic reactions in the mammal (MacLean, 1954).

There is increasing evidence of a sensory viscerosomatic influx into the limbic system. In pathological involvements, particularly of the temporal lobe, discharges in this part of the brain may be associated with a wide variety of auras involving all the body senses as well as a great number of feeling and emotional states. Among the more purely sensory auras are crude olfactory sensations; alimentary symptoms including taste, thirst, hunger, nausea; and somatic sensations ranging from pains to paresthesia. Prolonged rhythmic responses can be elicited in the pyriform area by olfactory, gustatory, and painful stimulation (MacLean, 1954). MacLean notes that all three of these senses accent the quality and intensity of a stimulus rather than its spatial relationships.

I do not wish to give the impression that this system is essentially one for viscerosomatic sensation, for it is also

implicated in a variety of complex ef-
fects such as the Klüver-Bucy syn-
drome (1958), memory losses (Pen-
field, 1958), the lowering of rage
thresholds in some cases or increased
docility after ablation in others (Bard
& Mountcastle, 1947; Green, 1958;
Schreiner & Kling, 1953; Spiegel, Mil-
ler, & Oppenheimer, 1940), and a va-
riety of cranial self-stimulation effects
(Brady, 1958; Olds, 1958; Olds &
Milner, 1954). The details of these
relations are beyond the scope of the
present discourse. What I wish to
emphasize is that there are significant
sensory inputs into this system which
may relate to the hedonic and reinforc-
ing features of stimulation as compared
with either the cognitive or arousal
functions.

Stellar and his colleagues (Sprague,
Chambers, Stellar, Liu, & Robson,
1958; Stellar, Chambers, Liu, Levitt,
& Sprague 1958) at the University of
Pennsylvania have made some interest-
ing studies of chronic animals with ex-
tensive lesions in the lateral lemniscal,
primary sensory pathways. I was
struck with their description of the
lemniscal animals, i.e., the sensorially
restricted animals displayed little affect
in any situation, except, perhaps the
most extreme. They flex and tend to
pull away from a pinch, but do not
attack or show any autonomic response.
They showed little or no aversive reac-
tion to an ether cone, although they
lacrimated and sneezed. Prior to the
operation, they solicited petting and re-
sponded to it well; afterwards they gave
no reaction to petting.

These observations suggest that the
affective response to sensory stimuli
might be mediated by the lemniscal and
not the diffuse arousal system, which,
it might be noted, already seems to have
been theoretically overworked by psy-
chologists. I would like to remind you
of the "thalamic syndrome," described

by Head and Holmes (1911). Lesions
in the region of the thalamus are char-
acterized by "overreaction," by exces-
sive affectivity, for both pleasurable as
well as painful sensations. This is
pathology characterized by the affective
modification of sensation. Although the
"syndrome thalamique" is well known
to clinical neurologists, its exact mech-
anism is still not clear. For our pur-
poses, it is sufficient to note that in this
condition, the basic hedonic responses
to sensory stimulation, some of the
"Pleasures of Sensation" as well as
the displeasures, can be unmasked. In
short, I am suggesting that the affec-
tive consequences of sensory stimula-
tion are mediated by processes that
depend upon the primary projection
systems and their ramifications in
thalamic and old brain neural connec-
tions. Obviously much work is re-
quired to place these speculations upon
a more solid foundation.

SUMMARY

And now let me briefly summarize.
I alluded briefly to the different roles
of stimuli and sensory processes in the
behavioral economy of the organism.
I pointed out that traditionally sensory
processes were studied largely in rela-
tion to discrimination or so called cog-
nitive functions. But stimuli have been
shown to have other neural and behav-
ioral functions as arousal and reinforce-
ment. I then went on to discuss in
further detail certain of our own ex-
periments on taste as a model system to
show the relation of gustatory stimula-
tion to the control of consummatory re-
sponses and the reinforcement of in-
strumental behavior. It was shown
that there is increasing support for the
idea that gustatory stimulation per se is
capable of eliciting and reinforcing be-
havior in its own right. We might say
"sweet for sweet's sake." I then led

into the problem of hedonic or affective responses to such stimuli; and, although a good correlation between affective response and reinforcement could be demonstrated, a statement as to their exact relation must await further study, particularly at the physiological level. Finally, I speculated as to the psychophysiological mechanisms that might underlie reinforcement and the affective responses to sensory stimuli. My basic theme has been that sensory stimulation "qua stimulation" plays a significant role in the motivation as well as guidance of behavior—euphemistically we might say, in controlling behavior for the "Pleasures of Sensation."

REFERENCES

BARD, P., & MOUNTCASTLE, V. B. Some forebrain mechanisms involved in expression of rage with special reference to suppression of angry behavior. *ARNMD*, 1947, **27**, 362–404.

BARE, J. K. The specific hunger for sodium chloride in normal and adrenalectomized white rats. *J. comp. physiol. Psychol.*, 1949, **42**, 242–253.

BEEBE-CENTER, J. G. "Feeling and emotion." In H. Helson (Ed.), *Theoretical foundations of psychology*. New York: Van Nostrand, 1951. Pp. 254–317.

BINDRA, D. *Motivation, a systematic reinterpretation.* New York: Ronald, 1959.

BRADY, J. V. Temporal and emotional factors related to electrical self-stimulation of the limbic system. In, *Recticular formation of the brain*. Boston: Little, Brown, 1958. Pp. 689–704.

BREMER, F. Cerveau "isole" et physiologie due sommeil. *CR Soc. Biol., Paris*, 1935, **118**, 1235–1241.

BROWN, J. S. Pleasure-seeking behavior and the drive-reduction hypothesis. *Psychol. Rev.*, 1955, **62**, 169–179.

CAMERON, A. T. The taste sense and the relative sweetness of sugars and other substances. New York: Sugar Research Foundation, 1947. (*Sci. Rep. Ser.*, No. 9.)

CAMPBELL, B. A. Auditory and aversion thresholds of rats for bands of noise. *Science*, 1957, **125**, 596–597.

CAMPBELL, B. A. Absolute and relative sucrose preference thresholds for hungry and satiated rats. *J. comp. physiol. Psychol.*, 1958, **51**, 795–800.

CARPENTER, J. A. Species differences in taste preferences. *J. comp. physiol. Psychol.*, 1956, **49**, 139–144.

COLLIER, G., & SISKEL, M., JR. Performance as a joint function of amount of reinforcement and interreinforcement interval. *J. exp. Psychol.*, 1959, **57**, 115–120.

COPPOCK, H. W., & CHAMBERS, R. M. Reinforcement of position preference by automatic intravenous injection of glucose. *J. comp. physiol. Psychol.*, 1954, **47**, 355–357.

ENGEL, R. Experimentelle Untersuchungen über die Abhängigkeit der Lust und Unlust von der Reizstärke beim Geschmackssinn. *Arch. ges. Psychol.*, 1928, **64**, 1–36.

FRINGS, H. Biological backgrounds of the "sweet tooth." *Turtox News*, 1946, **24**, No. 8.

GALAMBOS, R. Suppression of auditory nerve activity by stimulation of efferent fibers to cochlea. *J. Neurophysiol.*, 1956, **19**, 424–437.

GREEN, J. D. The rhinencephalon and behavior. In, *Neurological basis of behavior*. Ciba Foundation Symposium. Boston: Little, Brown, 1958. Pp. 222–235.

GUTTMAN, N. Operant conditioning, extinction and periodic reinforcement in relation to concentration of sucrose used as reinforcing agent. *J. exp. Psychol.*, 1953, **46**, 213–224.

GUTTMAN, N. Equal reinforcement values for sucrose and glucose solutions compared with equal sweetness values. *J. comp. physiol. Psychol.*, 1954, **47**, 358–361.

HAGBARTH, K. E., & KERR, D. I. B. Central influences on spinal afferent conduction. *J. Neurophysiol.*, 1954, **17**, 295–307.

HAGSTROM, E. C., & PFAFFMANN, C. The relative taste effectiveness of different sugars for the rat. *J. comp. physiol. Psychol.*, 1959, **52**, 259–262.

HARLOW, H. F. Motivations as a factor in the acquisition of new responses. In, *Current theory and research in motivation*. Lincoln: Univer. Nebraska Press, 1953. Pp. 24–49.

HEAD, H., & HOLMES, G. Sensory disturbances from cerebral lesions. *Brain*, 1911, **34**, 102–254.

HEBB, D. O. Alice in wonderland or psychology among the biological sciences. In H. F. Harlow & C. N. Woolsey (Eds.), *Biological and biochemical bases of behavior*. Madison: Univer. Wisconsin Press, 1958. Pp. 451–467. (a)

HEBB, D. O. *A textbook of psychology.* Philadelphia & London: Saunders, 1958. (b)

HERNANDEZ-PEON, R. Central mechanisms controlling conduction along central sensory pathways. *Acta. neurol. latino. Amer.*, 1955, **1**, 256–264.

HUGHES, L. H. Saccharine reinforcement in a T maze. *J. comp. physiol. Psychol.*, 1957, **50**, 431–435.

KELLER, F. S., & SCHOENFELD, W. N. *Principles of psychology.* New York: Appleton-Century-Crofts, 1950.

KLING, J. W., HOROWITZ, L., & DELHAGEN, J. E. Light as a positive reinforcer for rat responding. *Psychol. Rep.*, 1956, **2**, 337–340.

KLÜVER, H. The temporal lobe syndrome produced by bilateral ablations. In, *Neurological foundations of behavior.* Ciba Foundation Symposium. Boston: Little, Brown, 1958. Pp. 175–186.

LEMAGNEN, J. Le role de la receptivite gustative au chlorure de sodium dans le mecanisme de regulation de la prise d'eau chez le rat blanc. *J. Physiol. Path. gen.*, 1955, **47**, 405–418.

LINDSLEY, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Ch. 14.

MCCLEARY, R. A. Taste and postingestion factors in specific hunger behavior. *J. comp. physiol. Psychol.*, 1953, **46**, 411–421.

MACLEAN, P. D. The limbic system and its hippocampal formation in animals and their possible application to man. *J. Neurosurg.*, 1954, **11**, 29–44.

MACLEAN, P. D., HORWITZ, N. H., & ROBINSON, F. Olfactory-like responses in pyriform area to non-olfactory stimulation. *Yale J. Biol. Med.*, 1952, **25**, 159–172.

MACLEOD, S. A construction and attempted validation of sensory sweetness scales. *J. exp. Psychol.*, 1952, **44**, 316–323.

MAGOUN, H. W. *The waking brain.* Springfield, Ill.: Charles C Thomas, 1958.

MILLER, N. E. Experiments on motivation. *Science*, 1957, **126**, 1271–1278.

MILLER, N. E. Liberalization of basic S-R concepts: Extensions to conflict behavior, motivation, and social learning. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959.

MONTGOMERY, K. C. Explanatory behavior and its relation to spontaneous alternation in a series of maze exposures. *J. comp. physiol. Psychol.*, 1952, **45**, 50–57.

OLDS, J. Self-stimulation of the brain. *Science*, 1958, **127**, 315–324.

OLDS, J. S., & MILNER, P. Positive reinforcement produced by electrical stimulation of septal area and other regions of the rat brain. *J. comp. physiol. Psychol.*, 1954, **47**, 419–427.

PENFIELD, W. The role of the temporal cortex in recall of past experience and interpretation of the present. In, *Neurological foundations of behavior.* Ciba Foundation Symposium. Boston: Little, Brown, 1958. Pp. 149–174.

PFAFFMANN, C. Gustatory nerve impulses in rat, cat, and rabbit. *J. Neurophysiol.*, 1955, **18**, 429–440.

PFAFFMANN, C. Taste mechanisms in preference behavior. *Amer. J. clin. Nutr.*, 1957, **5**, 142–147.

PFAFFMANN, C. *Flavor research and food acceptance.* New York: Reinhold, 1958.

PFAFFMANN, C. The afferent code for sensory quality. *Amer. Psychologist*, 1959, **14**, 226–232.

PICK, H., & KARE, M. Certain aspects of taste preference in chickens and calves. Paper read at Eastern Psychological Association, Atlantic City, April, 1959.

PILGRIM, F. J. The components of food acceptance and their measurement. *Amer. J. clin. Nutr.*, 1957, **5**, 142–147.

RICHTER, C. P. Self-regulatory functions. *Harvey Lectures*, 1942, **38**, 63–103.

RICHTER, C. P., & CAMPBELL, K. H. Taste thresholds and taste preferences of rats for five common sugars. *J. Nutr.*, 1940, **20**, 31–46.

SCHLOSBERG, H. Three dimensions of emotion. *Psychol. Rev.*, 1954, **61**, 81–88.

SCHREINER, L., & KLING, A. Behavioral change following rhinencephalic injury in the cat. *J. Neurophysiol.*, 1953, **16**, 643–659.

SHEFFIELD, F. D., & ROBY, T. B. Reward value of a nonnutritive sweet taste. *J. comp. physiol. Psychol.*, 1950, **43**, 471–481.

SHEFFIELD, F. D., ROBY, T. B., & CAMPBELL. B. A. Drive reduction versus consummatory behavior as determinants of reinforcement. *J. comp. physiol. Psychol.*, 1954, **47**, 349–354.

SHERRINGTON, C. *The integrative action of the nervous system.* London: Constable, 1906.

SHUFORD, E. H., JR. Palatability and osmotic pressure of glucose and sucrose solutions as determinants of intake. *J. comp. physiol. Psychol.*, 1959, **52**, 150–153.

SKINNER, B. F. *Science and human behavior.* New York: Macmillan, 1953.

SMITH, M., & DUFFY, M. The effects of intragastric injection of various substances on subsequent bar pressing. *J. comp. physiol. Psychol.*, 1955, **48**, 387–391.

SMITH, M., & DUFFY, M. Evidence for a dual reinforcing effect of sugar. *J. comp. physiol. Psychol.*, 1957, **50**, 242–247.

SOULAIRAC, A. La physiologie d'un comportement: L'Appetit glucidique et sa regulation neuro-endocrinienne chez les rongeurs. *Bull. biol.*, 1947, **81**, 273–432.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven, Conn.: Yale Univer. Press, 1956.

SPIEGEL, E., MILLER, H., & OPPENHEIMER, J. Forebrain and rage reactions. *J. Neurophysiol.*, 1940, **3**, 538–548.

SPRAGUE, J. M., CHAMBERS, W. W., STELLAR, E., LIU, C. N., & ROBSON, K. Chronic reticular and lemniscal lesions in cats. *Fed. Proc.*, 1958, **17**, 154.

STELLAR, E., CHAMBERS, W. W., LIU, C. N., LEVITT, M., & SPRAGUE, J. M. Behavior of cats with chronic reticular and lemniscal lesions. *Fed. Proc.*, 1958, **17**, 156.

STELLAR, E., HYMAN, R., & SAMET, S. Gastric factors controlling water and salt solution drinking. *J. comp. physiol. Psychol.*, 1954, **47**, 220–226.

TROLAND, L. T. *The fundamentals of human motivation.* New York: Van Nostrand, 1928.

WARREN, R. P., & PFAFFMANN, C. Early experience and taste aversion. *J. comp. physiol. Psychol.*, 1958, **52**, 263–266.

YOUNG, P. T. *Motivation of behavior.* New York: Wiley, 1936.

YOUNG, P. T. Psychologic factors regulating the feeding process. *Amer. J. clin. Nutr.*, 1957, **5**, 154–161.

YOUNG, P. T. The role of affective processes in learning and motivation. *Psychol. Rev.*, 1959, **66**, 104–125.

YOUNG, P. T., & GREENE, J. T. Quantity of food ingested as a measure of relative acceptability. *J. comp. physiol. Psychol.*, 1953, **46**, 288–294.

YOUNG, P. T., & SHUFORD, E. H., JR. Intensity, duration, and repetition of hedonic processes as related to acquisition of motives. *J. comp. physiol. Psychol.*, 1954, **47**, 298–305.

YOUNG, P. T., & SHUFORD, E. H., JR. Quantitative control of motivation through sucrose solutions of different concentrations. *J. comp. physiol. Psychol.*, 1955, **48**, 114–118.

THEORETICAL NOTES

THE CONCEPT OF STATISTICAL SIGNIFICANCE AND THE CONTROVERSY ABOUT ONE-TAILED TESTS

H. J. EYSENCK

University of London

Several controversial papers regarding the uses and abuses of the one-tailed test of significance have recently appeared (Burke 1953, 1954; Goldfried 1959; Hick 1953; Jones 1952, 1954; Marks 1951, 1953). As Goldfried (1959) points out, "the important question debated is not *if* it should be used, but rather *when* it should be used." It is suggested here that most of the disagreements emerging from this controversy stem from a misunderstanding of the term "significance," and it is further suggested that the same misunderstanding runs through many discussions of two-tailed tests as well. It will be suggested that in the sense in which Goldfried's statement is meant, it has been the wrong question which has been debated; neither one-tailed nor two-tailed tests should be used at all in the sense envisaged by most of the writers quoted.

The outcome of the statistical examination of experimental results is always stated in terms of the probability of disconfirmation of the null hypothesis; the set of values which these p values can take is continuous in the interval from 0 to 1. It is customary to take arbitrary p values, such as .05 and .01, and use them to dichotomize this continuum into a *significant* and an *insignificant* portion. This habit has no obvious advantage, if what is intended is merely a restatement of the probability values; these are already given in any case and are far more precise than a simple dichotomous statement. Indeed, gross absurdities result from taking these verbal statements too seriously; the difference between a C.R. of 1.90 and another of 2.00 is quite negligible, yet one falls on one side of the dichotomy, the other on the other side. This has led to such summary statements as: "almost significant," or "significant at the 10% level." If the verbal dichotomous scale is not satisfactory—as it clearly is not—the answer surely is to keep to the continuous p scale, rather than subdivide the verbal scale.

However, surplus meaning has accrued to the word "significant," and it has become a shibboleth which divides the successful from the unsuccessful research. It is frequently interpreted as almost guaranteeing reproducibility of results, while failure to reach significance is interpreted as disconfirmation. Hence the urgent desire to achieve respectability and significance by one-tailed tests, if need be, and the argument regarding when the cachet of "significance" can be bestowed upon a research result. Yet the argument, and the achievement or nonachievement of significance, do not alter the facts of the case, which are contained in the statement of the p value of the results. Anything beyond these facts depends upon interpretation, and is subjective; it does not alter the facts of the case in the slightest.

As an example of the necessity of interpretation, consider the a priori probability of the conclusion. Suppose that an experiment on ESP were carried out with all the precautions which human ingenuity can devise, so that even the most sceptical had to agree that no fault could be found with the experimental design. *Suppose* also that a p value of .05 were achieved. Would this be considered "significant," in the sense of guaranteeing reproducibility? Critics would point out quite rightly that where the a priori probability is very low, as in this case, much higher p values would be required to carry significance. Logicians are

agreed that interpretation of experimental results must call on all available knowledge about the subject in question; a priori probability is a kind of summary statement of much of this knowledge. It cannot be overlooked in arriving at a conclusion regarding "significance" when the term carries the surplus meaning indicated.

That interpretation comes into the problem very much is clear when we look at such conditions as those suggested by Kimmel (1957) as criteria for the use of one-tailed tests. He suggests, for instance, that they may be used if results in the opposite direction would be psychologically meaningless or could not be deduced from any psychological theory. These are obviously not objective criteria, but depend on what the author (or reader) considers psychologically meaningless, or the kind of theory he may hold. Opinions will differ, and consequently some readers will agree to the use of the one-tailed test in a particular case, others will not. Thus to some readers the results will appear *significant*, to others *insignificant*.

The whole argument seems to be about *words*, not about *facts:* Is the word "significant" to be used in a given situation, or is it not? This would only matter if the word carried some objective meaning not contained in the probability figures; we have argued that it does carry surplus meaning, but that this is not of an objective kind. Consequently, nothing important is changed by omitting the term altogether in the report, leaving interpretation to the reader. After all, the only true proof of reproducibility is reproduction! Verbal assertions of "significance" have no more meaning than the *droit du pour* at the court of Louis XIV.

The solution is to separate quite clearly and decisively the *objective statement of the probability of disproof* of the *null hypothesis* (by means of a two-tailed test), and the *subjective evaluation and interpretation of the results*. The reader would be able to accept the first statement as a statement of fact and would then be able to judge for himself the ar-

guments presented by the author regarding the *meaning* of these facts. These arguments might be based on results of previous experiments, predictions made on the basis of more or less widely accepted theories, number of cases involved, a priori lack of accepability of the conclusions, and other similar grounds; an explicit statement of the arguments would enable the reader to decide for himself the acceptability of the conclusions in a manner precluded by the simple statement of one-tailed probability. *A statement of one-tail probability is not a statement of fact, but of opinion, and should not be offered instead of, but only in addition to, the factual two-tailed probability;* if it is offered at all, it should be accompanied by a full statement of the arguments in favor of its facilitating a more meaningful interpretation of the data. In the writer's opinion, it would be better to drop such statements of one-tailed probability altogether and rely entirely on appropriate argumentation to establish the meaning of the observed (two-tailed) probabilities.

Implicit in this recommendation is the corollary that the mechanical evaluation of experimental results in terms of "significant" and "not significant" be dropped outright. Interpretation is implicit in the statement of one-tailed probabilities, but it is also implicit in the statement of two-tailed probabilities if these are *automatically* interpreted as being significant or not significant, with all the surplus meaning carried by these terms. The experimenter should give his (two-tailed) *p* values and then proceed to argue regarding the acceptability of the conclusions on the basis already indicated. There have appeared in the literature solemn discussions about the possible causes for discrepancies between two experiments, one of which gave significant, the other insignificant results; yet the respective *t* values were almost identical, one lying just on the one side, the other just on the other side, of the arbitrary 5% line. Such arguments are unrealistic and would be avoided if *p* values were compared, rather than verbal statements. Two experiments giving *p* values of .048

and .056 are in excellent agreement, although one is significant, while the other is not.

To summarize the main point of this note briefly, we would say that verbal statements regarding "significance" are at best supererogatory restatements in an inconvenient dichotomous form of results already properly stated in terms of a continuous system of p values; at worst they carry unjustified surplus meaning of an entirely subjective kind under the guise of an objective and mathematically meaningful statement. Subjective judgments of reproducibility cannot reasonably be based on the mechanical application of a rule of thumb whose only usefulness lies in the elementary instruction of undergraduates lacking in mathematical background; if they are to be made at all they demand complex consideration of a priori probabilities. It is suggested that the accurate and factual statement of probabilities (two-tailed) should be mandatory and that all subjective considerations, arguments, and judgments should be clearly separated from such factual statements. It is implied that judgments of "significance" belong with the subjective side, and it is also implied that the calculation of p values on the basis of one-tailed tests has no place in psychology.

REFERENCES

BURKE, C. J. A brief note on one-tailed tests. *Psychol. Bull.*, 1953, **50**, 384–387.

BURKE, C. J. Further remarks on one-tailed tests. *Psychol. Bull.*, 1954, **51**, 587–590.

GOLDFRIED, M. R. One tailed tests and "unexpected" results. *Psychol. Rev.*, 1959, **66**, 79–80.

HICK, W. E. A note on one-tailed and two-tailed tests. *Psychol. Rev.*, 1952, **59**, 316–318.

JONES, L. V. Tests of hypotheses: One-sided and two-sided alternatives. *Psychol. Bull.*, 1952, **49**, 43–46.

JONES, L. V. A rejoinder on one-tailed tests. *Psychol. Bull.*, 1954, **51**, 585–586.

KIMMEL, H. D. Three criteria for the use of one-tailed tests. *Psychol. Bull.*, 1957, **54**, 351–353.

MARKS, M. R. Two kinds of experiment distinguished in terms of statistical operations. *Psychol. Bull.*, 1951, **58**, 179–184.

MARKS, M. R. One- and two-tailed tests. *Psychol. Rev.*, 1953, **60**, 203–208.

A NOTE ON THEORY AND METHODOLOGY IN THE STUDY OF FIGURAL AFTEREFFECTS

R. ALLEN GARDNER

Quartermaster Research and Engineering Center, Natick, Massachusetts

Recently, Mathews and Wertheimer have derived a theorem from the Kohler-Wallach (1944) theory of figural after-effects in order to provide an experimental test to discriminate between that theory and the Osgood-Heyer (1952) theory. This test must fail, however, because the principle that is invoked is contained in both theories, and the same prediction is at least as readily derived from either. Instead, the contribution of Mathews and Wertheimer lies in drawing attention to some commonly overlooked, but logically necessary consequences of a satiation-type theory. The rule that displacements will be uniformly symmetrical with respect to the contours of any I-figure has come to be accepted as a general statement of Kohler's theory and empirical findings. For example, Sagara and Oyama (1957) reported several experiments in which asymmetrical displacements were found as a contradiction of Kohler and Wallach (1944). The interpretation that Kohler and Wallach gave for this same finding in their original paper is rejected by Sagara and Oyama as a special ad hoc suspension of the rule of symmetry. In part, the object of the following analysis is to show that asymmetrical displacements are a logically necessary consequence of Kohler's assumptions in a wide variety of situations, while the rule of symmetry only applies to a restricted class of figural aftereffects.

The satiation of Kohler and Wallach and the neural adaptation of Osgood and Heyer are processes which arise from the contours of a figure when it is seen against a contrasting field. *Satiation* spreads to regions about the contours of the figure in a manner determined by electrotonic field-forces assumed to exist in the visual projection area of the cortex. *Neural adaptation* spreads about the contours in a manner determined by the statistical spread of neural excitation in the visual cortex caused by "the simultaneous action of physiological nystagmus and reciprocal overlap of dendritic processes" (Osgood & Heyer, 1952, p. 102). As Osgood and Heyer freely admit, once each theory has accomplished its genuflections to the particular sort of ritualistic physiological speculation favored by its authors, the results are essentially equivalent. With time, a gradient of satiation (or neural adaptation) is formed about the contour. The gradient is highest directly over the contour and decreases with distance in either direction, tending to be sigmoid in shape due to an assumed flattening at the peak. In general, displacements will be from regions where the gradient is high to regions where it is low. The derivation up to this point is illustrated in Fig. 1, where CD represents a line presented as an I-figure, P and P' represent the loci of T-figures presented after CD has been removed, and the arrows indicate the direction of displacements. If either P or P' is brought too close to CD, the gradients will extend symmetrically in both directions so that no displacement can occur. The extent of displacement has been found to increase with distance from the contour up to a maximum and then decrease again (the distance paradox). To account for this, both theories assume that the extent of displacement is a complex function of the height and the steepness of the gradients. Such symmetry of direction and extent, with CD representing the axis of symmetry, can be expected only as long as the I- and T-figures are analogous to the diagram of Fig. 1 which will be referred to here as Paradigm I. The analogy will continue to hold when the I-line, CD, is widened to become an oblong or square, but only so long as the T-figure does not become appreciably smaller than the I-figure.

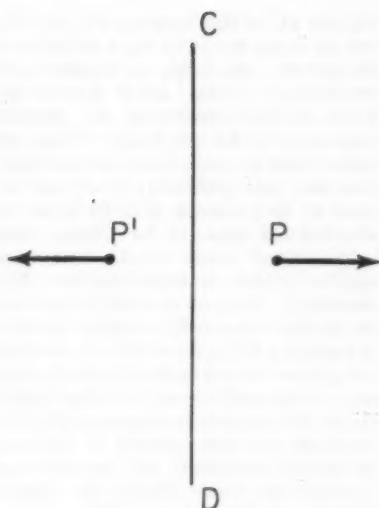Paradigm I is the illustration most

C

P' P

FIG. 1. Paradigm I.

frequently mentioned in the theoretical discussions of both Kohler and Wallach and Osgood and Heyer. Also, analogies of Paradigm I comprise the more intuitively obvious and striking of the various demonstrations reported by Kohler and Wallach, as well as those most reliably confirmed by subsequent investigators.[1] Thus, it is understandable that this simple paradigm has come to be accepted as the general model of figural aftereffects according to satiation theory. As Kohler would say, "Nothing could be farther from the truth." Consider the diagram in Fig. 2. Here the square ABCD represents a solid I-figure presented against a contrasting field (e.g., black square on white field). The point P represents the locus of a T-figure appreciably smaller than ABCD that is presented outside the contours of the I-figure and near the contour CD, while Q represents the locus of a similar T-figure presented near CD but within the contours of the I-figure. The solid ar-

[1] Extensive reviews of this literature have been published recently by McEwen (1958), Sagara and Oyama (1957), and Spitz (1958).

rows represent the symmetrical displacements on either side of CD that would be predicted merely by superimposing the I-line of Paradigm I on the new diagram. Clearly, this would not be justified. Unless the I-figure is so wide or the I-period is so short that the tissue satiating currents flowing from the opposite contour are negligible, displacement at Q should not be symmetrical with displacement at P. Without doing violence to the rest of the theory it is necessary to suppose that the satiating currents arising from the opposite contours, AB and CD, combine with each other in many cases. As a result, the satiation at loci inside the contours will always be greater than that at an equal distance outside the contours. At first, the cross sectional pattern of satiation will be saddle-shaped with the level of satiation in the cleft of the saddle rising more rapidly than that of the outward sloping gradients. With time a relatively flat dome-shaped pattern will emerge which spans the distance between the contours. Eventually, a homogeneous degree of satiation will be obtained between the contours and extending for some distance beyond them (Kohler & Wallach, 1944, p. 343–344). A precise statement of the rate at which each stage will appear and then give way to the next is beyond the scope of the theory in its present form. Nevertheless, with a suitable choice of I-figure, T-figure, and I-period, it should be possible to find a locus Q between the area of greatest satiation and the contour CD such that the direction of displacement will be *towards*
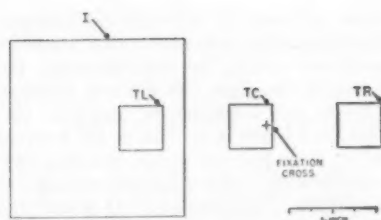

FIG. 2. Paradigm II.

FIG. 3. I- and T-figures used by Mathews and Wertheimer (1958).

CD. This is indicated by the broken arrow in Fig. 2.

Despite the claim of Mathews and Wertheimer (1958), which they have documented with a personal communication from Osgood, precisely the same theorem can be derived from the Osgood-Heyer theory. Osgood and Heyer (1952, p. 108) explicitly state that the statistical distributions of neural adaptation arising from opposite contours must overlap and summate with each other in a manner analogous to the satiation of Kohler and Wallach. Their Fig. 2 (1952, p. 105) is a cross sectional representation of Paradigm I. It becomes Paradigm II merely by superimposing upon it the overlapping gradient of neural adaptation which might be expected from a second contour. If this is done, the predictions of the Mathews-Wertheimer theorem must follow. Of course, to obtain a prediction for a specific I-figure it will no longer be possible to avoid specifying the parameters of the statistical distributions as it was in the original analysis presented by Osgood and Heyer.

Mathews and Wertheimer (1958) performed an experiment which they designed as a test of their theorem. The I- and T-figures they used are drawn to scale from their description in Fig. 3. The I-figure was a solid, black, 2-in. square corresponding to ABCD in Fig. 2, while TL, TC, and TR were three ½-in. squares presented simultaneously as a T-figure after a 2-min. inspection of the I-figure. The locus of TL corresponds to locus Q in Paradigm II. Each of 16 observers (Os) was asked to report which distance was greater, that between

TL and TC or that between TC and TR, and 14 chose the right hand distance as the greater. According to Mathews and Wertheimer: "This result favors the Kohler-Wallach prediction, by bimodal expansion, at the 1% level." Since no further data are submitted, we must suppose that this probability level was arrived at by assuming that 16 other Os who had not inspected the I-figure—that is, a control group—would have been equally divided in their choices. Unfortunately, those of us who do not concur in this admittedly common practice of assuming the performance of a control group must remain in doubt as to the outcome of the experiment. Further elaboration of this point is unnecessary as the importance of such a control in the study of figural aftereffects has already been discussed at some length by Brown (1953) and Wertheimer (1954).

In any event, even with a control group, the particular T-figure chosen by Mathews and Wertheimer could never be used to study the displacement of the square TL. All objects in the T-figure must be assumed to exhibit some displacement. Consequently, the best possible measurement of the extent or direction of displacement is the relative displacement of a near T-object when compared with that of a second, remote T-object. When the relative displacements of both the near and the remote T-objects are referred to that of an intermediate, third T-object no meaningful result can be obtained. The Os could report that the distance TC—TR is greater than the distance TL—TC when both TL and TC are displaced *away* from the contour in any case where the displacement of TR is greater than the combined displacement of TL and TC. Indeed, this might occur very commonly due to the distance paradox. Similarly, the distance TL—TC could be reported as greater even when TL is displaced *towards* the contour in any case where the displacement of TC away from the contour is greater than the combined displacement of TL and TR. The ubiquitous distance paradox would readily account for this outcome also.

Fortunately, Fox (1951) has already reported two technically more adequate experiments that fulfill the topological requirements of the Mathews-Wertheimer theorem. In Experiment I, Fox found that T-figures presented on the contour or just inside of it were displaced outwards, but to a smaller extent than T-figures presented outside the contour. Using somewhat different I-figures, in Experiment II he found that displacements were symmetrically away from the nearest contour for a variety of loci on either side of the contour. Fox concluded in favor of Kohler and Wallach on the basis of Experiment II, attributing the asymmetrical results of Experiment I to an artifact not present in Experiment II. It might at first appear that in order to preserve the Mathews-Wertheimer theorem, Fox should have based his conclusion on the results of Experiment I and explained away the symmetrical results of Experiment II. But this, too, would oversimplify the situation and underestimate the full power of the satiation theory. Clearly, both the initial pattern of satiation and its variation as a function of time must depend upon the particular I-figure that is in question. In the complex situation described by Paradigm II Fox's results can be accounted for simply by noting that the I-period he used must have been long enough to produce asymmetrical displacements with the I- and T-figures of Experiment I, but not with those of Experiment II.

The Mathews-Wertheimer theorem represents Case 1 of Paradigm II. It is also possible to derive its converse, Case 2, where the direction of displacement at loci outside the contours reverses with time under certain conditions. That this case, too, is implied as a consequence of satiation theory was noted by Kohler and Wallach (1944):

Suppose that the T-squares lie half inside and half outside the area of the I-oblong [on opposite contours]. If the inspection period is not too short the distance between these squares will decrease. Even with T-squares which lie entirely outside the oblong but just adjacent to it the same will be observed —always provided that the inspection period

has been sufficiently long . . . the direction of the effect depends upon the degree in which the medium has been altered by previous inspection. Also, during late stages of the inspection period a figure process may not affect the medium in quite the same way as it does in the beginning. Incidentally, in the same arrangement slight variations of the I-object may turn the effect in one direction or the other (p. 297).

The existence of Case 2 displacements towards the contour that Kohler and Wallach reported have since been confirmed by Smith (1954) and by several Japanese studies reviewed by Sagara and Oyama (1957). The explanation seems to depend upon the existence of I-figures which give rise to gradients which are steeper inside the contour than outside the contour. Both Kohler and Wallach (1944, pp. 339–340) and Osgood and Heyer (1952, p. 108) invoke this principle to account for an otherwise anomalous discrepancy between outline and solid figures. In such a case the peak will be shifted outwards somewhat, and further summation and interaction with the gradient extending from the opposite contour should increase this shift. The effect will not always be observed, of course; but with a suitable choice of I-figure, T-figure, and I-period, it should be possible to find a locus p outside the contour such that displacement will be towards the contour. In connection with this, Sagara and Oyama (1957) report that where Case 2 reversals were found the extent of nonreversal displacement was greater inside the contours than outside. This is precisely the result that would be expected if the inside gradient is steeper than the outside gradient in Case 2.

There is essentially no contradiction between the predictions of Case 1 and those of Case 2. Each is appropriate to a separate class of I-figures. It is intuitively obvious that different classes of I-figures must give rise to different patterns of electrical activity in the visual cortex. To claim otherwise is to do violence to language. However, the various patterns are related by a common neurological mechanism. According to Kohler

and Wallach, this mechanism is the electrotonic field-forces of satiation. According to Osgood and Heyer, it is the statistical distribution of neural adaptation. Neurological evidence has been cited on both sides. To be sure, the neurological data available to the authors were not sufficient to permit either theory to specify the particular pattern of cortical activity that would be produced by any particular set of conditions. This is understandable when we consider that the evidence that is cited pertains to phenomena that take place within fractions of a second, while figural aftereffects can extend over several minutes at least. But, with disarming frankness, both Kohler and Wallach and Osgood and Heyer admit the need for much further research in this area. In the interim, the gap is bridged by boldly assuming that in cases of Paradigm I the long-term neurological processes will be analogous to the short-term processes, while in cases of Paradigm II summation and interaction over time will produce precisely the gradients that are necessary to account for the figural aftereffects that occur. Thus, whether the neurophysiologists of the future opt for field-forces, or statistically independent units, or—as in modern physics—some happy combination of the two, it is probably safe to say now that perceptual phenomena are related to neural activity.

REFERENCES

Brown, K. T. Methodology for studying figural after-effects and practice effects in the Muller-Lyer illusion. *Amer. J. Psychol.*, 1953, **66**, 629–634.

Fox, B. H. Figural after-effects: Satiation and adaptation. *J. exp. Psychol.*, 1951, **42**, 317–326.

Kohler, W., & Wallach, H. Figural after-effects: An investigation of visual processes. *Proc. Amer. Phil. Soc.*, 1944, **88**, 269–357.

McEwen, P. Figural after-effects. *Brit. J. Psychol.*, *Monogr. Suppl.*, 1958, **31**, 1–106.

Mathews, T., & Wertheimer, M. An experimental test of the Kohler-Wallach and the Osgood-Heyer theories of figural after-effects. *Amer. J. Psychol.*, 1958, **71**, 611–612.

Osgood, C. E., & Heyer, A. W., Jr. A new interpretation of figural after-effects. *Psychol. Rev.*, 1952, **59**, 98–118.

Sagara, M., & Oyama, T. Experimental studies on figural after-effects in Japan. *Psychol. Bull.*, 1957, **54**, 327–338.

Smith, K. R. "Attraction" in figural after-effects. *Amer. J. Psychol.*, 1954, **67**, 174–176.

Spitz, H. H. The present status of the Kohler-Wallach theory of satiation. *Psychol. Bull.*, 1958, **55**, 1–28.

Wertheimer, M. Constant errors in the measurement of figural after-effects. *Amer. J. Psychol.*, 1954, **67**, 543–546.

SCALE CLASSIFICATION AND STATISTICS

JOHN GAITO

Wilkes College

The classification of psychological scales into nominal, ordinal, interval, and ratio categories by Stevens (1951) has been a significant contribution to psychophysical and measurement theory. This classification has been widely accepted and applied by psychologists in conversation, teaching, and in publication (e.g., Edwards, 1958; Guilford, 1954; Senders, 1958; Siegel, 1956). In considering these scales Stevens has specified the appropriate statistical measures for use with each. This latter aspect has had a tremendous impact on writers of statistical texts. For example, these categories form the basis for the texts by Senders (1958) and by Siegel (1956). There is no doubt that this classification has great practical and theoretical value. However, we feel that the application of this classification has led to some misunderstanding in regard to the use of various statistical techniques and to an overemphasis on the utility of nonparametric techniques in psychological research.

As an example of this tendency Siegel (1956, p. 26) maintains that nonparametric tests of significance should be used with subinterval type data. He lists the requirement of interval scale as one of the assumptions for the use of the analysis of variance. However, this assumption cannot be found if one looks to the mathematical bases of assumptions (Eisenhart, 1947). In fact a noted statistician, Kempthorne (1955), after showing mathematically that the normal theory analysis of variance test can approximate the randomization test, states that

This serves as some theoretical basis for the fact which has been noticed by most statisticians, that the level of significance of the analysis of variance test for differences between treatments is *little affected by the choice of a scale of measurement for analysis* [italics added].

The important consideration for the use of the analysis of variance is not that the data have certain scale properties but that *the data can be related to the normal distribution*, plus approximating the other assumptions of independence and homogeneity of errors. Furthermore, it must be emphasized that the normal distribution refers to the distribution of those portions of the data which are used as the appropriate estimates of error and not necessarily to the distribution of the observations. This is an important consideration because in multivariable designs the variation within each group may not distribute normally but the portions of certain interaction terms may. In many of these designs some of the interaction effects provide the valid estimate of error in testing certain effects.

It would seem that the comments concerning the relation between the various scales and the appropriate statistical procedures should serve merely as a guide. In some cases empirical data may indicate that rigid adherence to such statistical methods is not required and wasteful of the data. The author has propounded elsewhere this viewpoint and cited relevant data concerning the analysis of variance technique (Gaito, 1959).

Furthermore, the same data may be considered to have the properties of two or more scales, depending on the context in which it is considered. For example, if we look at the response of one subject (S) to a single item, the properties of the data are those of a nominal scale, i.e., right or wrong. However, if we concentrate on the total score for one S or the total scores for a group of Ss, we have at least an ordinal scale. This is similar to the situation of approximating the binomial distribution (a distribution having nominal, or possibly ordinal, scale properties) by the normal distribution (a distribution having at least ordinal, pos-

sibly interval, scale properties) when n increases and p is close to .5. Another example is the relation between the sign test and the binomial distribution test. For a given set of data these two methods will give the same result; yet Siegel lists the former as of an ordinal nature, the latter as nominal data.

It appears that in dealing with descriptive statistics the comments made about the use of nonparametric statistics for subinterval-type data are most appropriate. However, in dealing with tests of significance such statements are not as appropriate, even though descriptive statistics are involved in the procedure. For each test the mathematical requirements are expressed in the assumptions, and nothing concerning specific scale properties is required. Supplementing the requirements will be evidence to indicate the effects on the data of failure to meet the assumptions (Gaito, 1959; Lindquist, 1953). Thus if empirical research indicates little effect on the data when assumptions are not met or merely approximated, the investigator may use a parametric technique and allow for the effects of the deviations from assumptions in his interpretation.

It is encouraging to note that some individuals have been reluctant to embrace wholeheartedly the nonparametric technique (e.g., Guilford, 1954; McNemar, 1957, 1958; Savage, 1957). In a review of the statistical literature in the *Annual Review of Psychology,* Grant (1959, p. 137) appropriately summarizes this position by the comment: *"Some much-needed negative thinking has recently appeared on nonparametric techniques"* [italics added].

In conclusion we wish to restate that the statements concerning scale properties and statistical procedures should be guides and that context, mathematical assumptions of statistical procedures, and the results of research concerned with failure to satisfy assumptions should be the ultimate determiners of the choice of statistical techniques.

REFERENCES

EDWARDS, A. E. *Statistical analysis.* New York: Rinehart, 1958.

EISENHART, C. The assumptions underlying the analysis of variance. *Biometrics,* 1947, **3,** 1–21.

GAITO, J. Nonparametric methods in psychological research. *Psychol. Rep.,* 1959, **5,** 115–125.

GUILFORD, J. P. *Psychometric methods.* New York: Wiley, 1954.

KEMPTHORNE, O. The randomization theory of experimental inference. *J. Amer. Statist. Ass.,* 1955, **50,** 946–967.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.

MCNEMAR, Q. On Wilson's distribution-free test of analysis of variance hypotheses. *Psychol. Bull.,* 1957, **54,** 361–362.

MCNEMAR, Q. More on the Wilson test. *Psychol. Bull.,* 1958, **55,** 334–335.

SAVAGE, I. R. Nonparametric statistics. *J. Amer. Statist. Ass.,* 1957, **52,** 331–344.

SENDERS, V. L. *Measurement and statistics.* New York: Oxford Univer. Press, 1958.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

STEVENS, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.